

Grado en Estadística

Título: Más allá de la brecha salarial

Autor: Alicia León Molina

Director: Ernest Pons Fanals

Departamento: Departament d'Econometria, Estadística i Economia Espanyola

Convocatoria: Junio 2019



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística

Resumen y palabras clave

Existe un elevado interés en la actualidad para identificar cuáles son los motivos que justifican una persistencia tan elevada de diferencias salariales entre hombres y mujeres. En este sentido, son muchos los ámbitos de actividad en los que las discriminaciones van desapareciendo; no obstante, la brecha salarial continua. Existen visiones alternativas sobre las razones de esta discriminación, pero en este trabajo se pretende explorar la posibilidad de que dichas diferencias pueden estar relacionadas (al menos en parte) con diferentes actitudes y/o estrategias en relación con la formación; incluyéndose dentro del concepto "formación" la decisión de estudiar o no, el nivel de estudios adquirido, la formación continuada o no reglada o la inserción laboral. Para ello se analizarán tres fuentes de información: dos bases de datos, una sobre la participación en la educación de los individuos españoles y otra sobre la inserción laboral de los estudiantes universitarios; esto complementado con una exploración de concienciación en torno al contexto de brecha salarial, inserción laboral o igualdad.

Palabras clave: diferencias de género, estudios, inserción laboral, Twitter, Google Trends, análisis estadístico, minería de datos, reglas de asociación, clustering jerárquico

Abstract and key words

Currently exists an increasing interest to identify the possible reasons that could justify such a high persistence of wage differences between men and women. In this sense, there are many areas of activity in which discriminations are disappearing. However, the pay gap continues. There are alternative point of views on the reasons for this discrimination, but this paper aims to explore the possibility that these differences may be related (at least in part) to different attitudes and / or strategies in relation to the formation each individual has. Including within the concept "formation" the decision to study or not, the level of studies acquired, the continuous or non-regulated training or the labour insertion. To do so, three sources of information will be analyzed: two databases, one on the participation in the education of Spanish individuals and another on the labour insertion of university students. This is complemented by an exploration of awareness around the context of the pay gap, labour insertion or equality.

Key words: gender differences, studies, labour insertion, Twitter, Google Trends, statistical analysis, data mining, association rules, hierarchical clustering

Clasificación AMS

62-01 Data Analysis

62-09 Graphical methods

62H15 Hypothesis methods

62D05 Sampling theory, sample surveys

62H30 Classification and discrimination; cluster analysis

62P25 Applications to social sciences

ÍNDICE

INTRODUCCIÓN	1
METODOLOGÍA	6
DOCUMENTACIÓN	10
CUERPO DEL TRABAJO	13
CAPÍTULO I. ¿BRECHA EDUCACIONAL?	13
1.1. <i>EDUCACIÓN FORMAL</i>	14
1.2. <i>EDUCACIÓN NO FORMAL</i>	35
1.3. <i>APRENDIZAJE INFORMAL</i>	47
CAPÍTULO II. ¿BRECHA EN INSERCIÓN LABORAL?	50
2.1. <i>DURANTE LA FORMACIÓN</i>	52
2.2. <i>INCORPORACIÓN AL MERCADO LABORAL</i>	54
2.3. <i>PRIMER EMPLEO</i>	57
2.4. <i>EMPLEO ACTUAL</i>	59
CAPÍTULO III. ANÁLISIS DE CONCIENCIACIÓN	65
3.1. <i>TWITTER</i>	65
3.2. <i>GOOGLE TRENDS</i>	68
CONCLUSIONES	76
POLÍTICAS SOCIALES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	77
BIBLIOGRAFÍA	80
ANEXOS	82
PREPROCESSING	82
<i>FORMACIÓN</i>	82
<i>INSERCIÓN LABORAL</i>	82
GRÁFICOS	85
RESULTADOS R	85
SCRIPTS	87
ENCUESTAS	154

Introducción

Históricamente nos encontramos en una sociedad donde hay diferencias de género en diferentes ámbitos, aunque estas han ido evolucionando a lo largo de los años. Si bien es cierto que actualmente (en España) las mujeres tienen, legalmente, los mismos derechos que los hombres, aún existen diferencias a veces perceptibles, que hacen que mujeres y hombres no sean iguales ante la vida.

Las mujeres representan aproximadamente el 50% del capital humano de un país y de su talento. Y, por lo tanto, desde un punto de vista económico resultan un activo fundamental para contribuir al progreso económico social.

Esto, dejando de lado aspectos éticos y sociales que no son menos importantes. Pero incluso desde un punto de vista de recursos, es importante encontrar la manera de aprovechar este talento. Ahora bien, durante muchos años la contribución de las mujeres al mercado del trabajo ha estado inferior a la de los hombres, por muchas razones que no serán analizadas en este trabajo.

En cualquier caso, partimos de la constatación de que a día de hoy hay una relación asimétrica entre el género y todo aquello que tiene que ver con el mercado de trabajo.

El punto de partida es constatar estas diferencias. Por un lado, hay diferencias en la "participación" en el mercado laboral. Pero estas han demostrado en los últimos años una tendencia a disminuir.

Nota: En 1985 solo participaba en el mercado laboral el 35% de las mujeres en edad de trabajar (15-64), en la actualidad la participación de las mujeres en el mercado laboral español se sitúa por encima de la media europea con el 68%. A pesar de este avance, y de estar cerca de la media de los países industrializados, todavía estamos lejos de conseguir un mercado de trabajo que garantice la igualdad de oportunidades entre hombres y mujeres.

En cambio, la diferencia en cuanto a los salarios parece que está mucho más "enquistada". Si consultamos la *Encuesta de Estructura Salarial* del INE, podemos comprobar que entre 2008 y 2016 estas diferencias se han mantenido prácticamente constantes, en un valor global de casi el 30%. A continuación, se puede ver en la siguiente tabla como estas diferencias son crecientes con la edad.

Tabla I-1. Diferencias entre la ganancia anual media de mujeres y hombres

	2008	2009	2010	2011	2012	2013	2014	2015	2016
De 20 a 24 años	26.1%	22.3%	24.3%	23.6%	23.2%	26.0%	23.3%	31.1%	29.7%
De 25 a 29 años	17.1%	14.6%	17.6%	18.5%	17.2%	16.5%	17,4%	16.3%	14.3%
De 30 a 34 años	22.4%	22.8%	20.3%	22.9%	25.2%	23.1%	18,9%	21.5%	21.7%
De 35 a 39 años	26.8%	25.9%	22.8%	26.0%	28.1%	28.3%	24,9%	24.3%	24.9%
De 40 a 44 años	29.5%	30.1%	30.1%	28.5%	28.9%	30.2%	29,2%	27.2%	25.8%
De 45 a 49 años	29.6%	30.6%	34.0%	32.0%	35.7%	35.2%	36,4%	32.3%	30.9%
De 50 a 54 años	29.6%	34.7%	36.9%	35.9%	36.0%	37.6%	34,5%	33.8%	33.4%
De 55 a 59 años	29.7%	32.7%	39.4%	37.5%	40.0%	40.5%	39,9%	38.5%	36.8%
De 60 a 64 años	24.4%	20.5%	20.4%	17.6%	19.2%	17.9%	30,3%	30.8%	30.3%
TOTAL	28.0%	28.2%	29.1%	29.8%	31.5%	31.6%	30.3%	29.6%	28.8%

Fuente: Encuesta de Estructura Salarial (INE)

Ahora bien, la comparación directa proporciona una visión “sesgada”, porque las diferencias podrían ser debidas a otros factores ajenos al género: nivel de estudios, sector, cargo, etc.

En un artículo [1] reciente se comprueba como se ha producido un avance importante en los últimos años en lo que respecta a las diferencias salariales entre hombre y mujeres.

Controlando por todas las características observables, la brecha salarial entre hombre y mujeres se ha reducido un 33% desde 2002. No obstante, todavía nos encontramos lejos de la igualdad de género en materia salarial y se observan algunas dinámicas preocupantes.

Los principales resultados del trabajo son los siguientes. En primer lugar, los autores encuentran que la brecha por hora trabajada, diferenciando solo por características socioeconómicas, es del 17%, mientras que, cuando además se recogen las características del lugar de trabajo y de la empresa, esta se reduce hasta el 13%. En segundo lugar, el estudio concluye que las brechas ajustadas aumentan con la edad, se reducen con el nivel educativo y aumentan con la antigüedad en la empresa. Las brechas ajustadas también son más altas en los contratos indefinidos y a tiempo completo y en los sectores y ocupaciones con una mayor presencia masculina. Finalmente, la brecha salarial ajustada es especialmente alta en la parte superior de la distribución de los salarios, evidencia del *techo de cristal*¹ en el mercado laboral español y, sorprendentemente, tiende a ser mayor en las grandes empresas que en las pequeñas.

¹ Se denomina techo de cristal a la limitación velada por el ascenso laboral de las mujeres al interior de las organizaciones. Se trata de un techo que limita sus carreras profesionales.

En este artículo los autores concluyen que sus resultados apuntan a la necesidad de reforzar políticas públicas y de empresa que busquen fomentar la igualdad de género en materia salarial y, más generalmente, la igualdad de oportunidades en el mercado laboral.

Entre estas políticas estarían las actuaciones destinadas a mejorar la conciliación entre la vida familiar y vida laboral y a fomentar que los/las dos integrantes de la pareja se hagan corresponsables. Medidas en este sentido podrían ser las de universalizar la educación de 0 a 3 años, ampliar los permisos de paternidad y fomentar la flexibilidad de la ocupación. Además, aquellas políticas que ayuden a evitar comportamientos discriminatorios y las que contribuyan a corregir el problema de infrarrepresentación femenina en los puestos de liderazgo, incluyendo la introducción de cuotas de género progresivas y temporales.

Este trabajo parte de la brecha salarial existente, la cual se conoce que existe y ha sido estudiada y analizada por profesionales del sector, y se quiere centrar en analizar los posibles factores que se puedan producir a lo largo de la “carrera profesional” de una persona en torno a la brecha salarial. Se pretende complementar la constatación de la brecha salarial con una búsqueda de cuáles son los factores que pueden estar incidiendo más directamente en esta brecha, con la finalidad de entender mejor el entorno contextual y analizando si hay más puntos diferenciadores entre hombres y mujeres.

Por un lado, se estudiará la educación y formación de las personas. Se explorarán variables concretas en torno a este tema, con la intención de poder detectar, o no, posibles *brechas de género* en materia educativa. Esto se complementará estudiando, de la misma manera, la inserción laboral para ampliar la visión de la carrera profesional. Este punto se centrará únicamente en los universitarios, siendo en los máximos niveles educativos donde se podría o debería esperar una menor aparición de diferencias entre géneros.

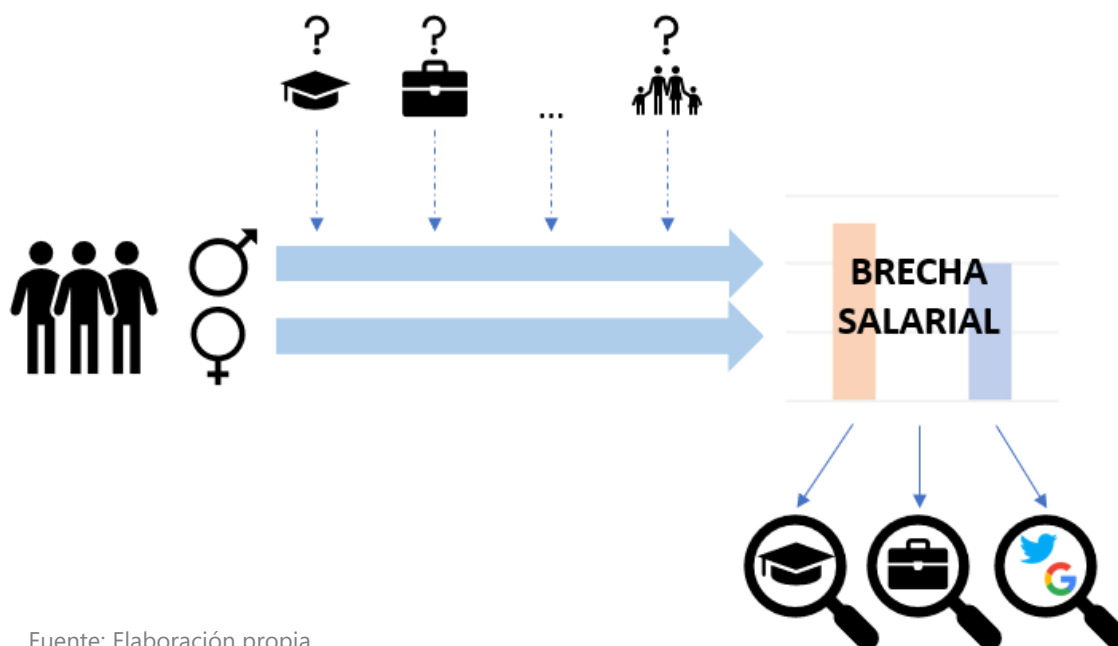
Con la ayuda de la *Encuesta sobre la Participación de la Población Adulta en las Actividades de Aprendizaje (EADA 2016)* realizada por el Instituto Nacional de Estadística (INE), se estudiará el comportamiento y el perfil de los adultos españoles en torno a la formación educativa, buscando posibles diferencias entre géneros. Posteriormente, mediante la *Encuesta de inserción laboral de titulados universitarios*, también del INE, se buscará analizar el comportamiento y las vivencias de hombres y mujeres en el paso de la vida educativa a la vida laboral, buscando detectar posibles diferencias.

Sobre estas dos bases de datos, se tratará de emplear distintas metodologías, siempre con el objetivo final de poder detectar diferencias, si las hay. En el caso de la educación se hará un análisis más "clásico", analizando los gráficos cruzados y sometiendo las variables a contrastes de hipótesis, todo complementado con la construcción de clústers; y el aspecto de la inserción laboral, se abordará mediante reglas de asociación, también complementado con análisis clásico cuando sea pertinente.

Por otro lado, y con un enfoque muy distinto, se explorará mediante Twitter y Google la concienciación en torno a estas problemáticas. Con la motivación de abordar un tema social uniendo metodologías distintas, esta parte dará un punto de visión sobre la concienciación o preocupación de las personas en torno a estos temas, un aspecto crucial para la erradicación o mejora de posibles problemáticas, en este caso, de desigualdad entre géneros en aspectos educativos y de inserción en el mundo laboral. Aunque la investigación principal se centra en los datos del INE, con esta parte final se busca abordar el tema de diferentes maneras, lo que podría permitir un mejor conocimiento del entorno y el contexto social.

A continuación, se presenta un esquema representativo del trabajo:

Ilustración I-1. Esquema general del trabajo



Fuente: Elaboración propia

Destacar que este trabajo no busca encontrar las causas directas de la brecha salarial, sino estudiar, de manera exploratoria, posibles factores que la propicien o que intervengan en el contexto y ver así si hay puntos clave en la carrera profesional de una persona que puedan contribuir a la aparición de esta o que puedan explicarla mejor.

Esta iniciativa surge, como ya se ha ido introduciendo, del hecho de que, aunque con los años parece que mujeres y hombres se acercan a la igualdad en ámbitos como la educación y las oportunidades laborales, esto no se ve reflejado en una disminución de la brecha salarial, o por lo menos, no se observa una reducción proporcional. Es por eso, que se han decidido estos momentos en la vida profesional de una persona, para ser sometidos a un análisis estadístico y poder contribuir, en la medida de lo posible, al entendimiento de qué está ocurriendo en torno al gran problema social que es la brecha salarial y, en general, las desigualdades de género.

Antes de emprender el desarrollo del trabajo, expresar mi más sincero agradecimiento al Doctor Ernest Pons por su apoyo, sus ideas y su implicación. En todo momento he sentido su estrecha colaboración, algo que ha favorecido tanto a mi motivación incluso en los momentos más inciertos, como al rumbo del trabajo con el cual estoy muy satisfecha, independientemente de los resultados.

También agradecer a Roger, mi mayor fuente de inspiración, sin su incansable apoyo, motivación y paciencia, este trabajo no sería lo que es. A mi familia por estar ahí, incluso cuando han estado meses sin verme en las etapas más intensas de trabajo. A mis compañeros becarios, por entendernos mejor que nadie, aprovechando cada minuto para avanzar y llevarlo todo para delante. Finalmente, a amigos/as y compañeros/as por haber compartido los buenos y los malos momentos de principio a fin.

Con esto, presento este trabajo como el punto de partida de mi carrera profesional, la cual emprendo con muchísimas ganas, expectante por todo lo que está por venir.

Metodología

En este apartado, se explicará la metodología empleada, las técnicas estadísticas utilizadas, la procedencia de los datos y las explicaciones pertinentes de estos y los recursos informáticos utilizados. También se llevarán a cabo las explicaciones necesarias para el mejor entendimiento de los procesos llevados a cabo a lo largo del trabajo; así como puntualizaciones detalladas en cuanto a los métodos utilizados.

1. ¿Brecha educacional?

En el capítulo 1, se han empleado tests de χ^2 para la comprobación de independencia o dependencia de las variables en cuestión con el género de los individuos, como método de detección de aspectos diferenciales entre hombres y mujeres. Ha sido empleado esta prueba puesto que el Test exacto de Fisher tendría un coste computacional muy alto y, al trabajar con una gran cantidad de datos, los resultados serían muy parecidos a los obtenidos con el aproximado χ^2 . Cuando esta prueba se ha querido realizar para datos numéricos, se ha empleado el test t-student.

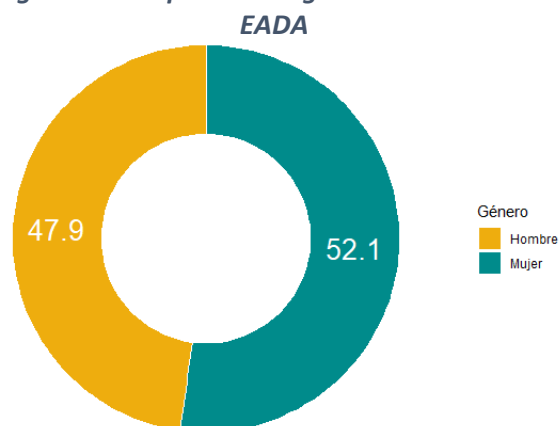
En las siguientes fuentes [2] [3] [4] se puede encontrar tanto el desarrollo como los fundamentos teóricos de los métodos empleados, en los que se basa este trabajo.

También en este apartado se han usado técnicas de minería de datos, como es el clustering jerárquico. El criterio de agregación utilizado ha sido el método de Ward. Este criterio es de carácter aglomerado, es decir, que se inicia con tantos grupos como individuos y se van agrupando hasta tener todos los individuos en el mismo grupo. Ward se basa en la minimización de una función objetivo, utilizándose en el trabajo una con intención de minimizar la variancia. Las distancias entre clústers se han calculado mediante la distancia de Gower al cuadrado, con la intención de tener en cuenta variables categóricas. A continuación, se presentan las referencias teóricas [5] [6] en las que se basa esta metodología.

El primer apartado del trabajo se centrará en estudiar la ya mencionada *Encuesta sobre la Participación de la Población Adulta en las Actividades de Aprendizaje* de 2016 que realiza el INE. Sobre la fuente de los datos, la encuesta es de periodicidad quinquenal y se realiza a personas de entre 18 a 64 años. Está realizada en todo el territorio nacional. La base de datos original se compone de 23019 registros y 247 variables.

Dado que en la muestra de la EADA hay un mayor número de mujeres que de hombres, aunque leve, los resultados gráficos han de ser interpretados con cautela, pues siempre podrá haber

Figura I-1. Proporción de géneros en los datos



Fuente: Elaboración propia

una diferenciación simplemente por el hecho de un mayor número absoluto. Al haber un 52.1% de mujeres y un 47.9% de hombres, en los gráficos que se observe la proporción de géneros dentro de cada categoría, habrá un margen de error por este motivo. Es por esto que, cuando sea interpretable visiblemente, se realizarán gráficos que representen, para cada género, su distribución en las variables. Y, por otro lado, cuando con estos no sea tan fácil su comprensión visual, se observará la proporción

dentro de cada categoría, siempre teniendo en cuenta este margen de error. De todas maneras, y para concluir, nunca se validará la existencia empírica de diferencias mediante los gráficos, sino que esto se hará mediante el contraste de hipótesis correspondiente.

2. ¿Brecha en inserción laboral?

En el capítulo 2, se ha recurrido a otra técnica de minería de datos, como son los métodos asociativos, en concreto reglas de asociación. Estas consisten en encontrar comportamientos concretos en los datos, con la intención de encontrar relaciones entre los diferentes ítems que las componen, es decir, saber que cuando pase un hecho X, también sucederá otro hecho Y. Con estas se encontrarán reglas de sucesos que describan patrones, en este caso, implicando a la variable género y poder relacionar hechos con hombres o mujeres, si procede. El algoritmo concreto que se utilizará será apriori. La metodología y teoría en cuanto a estos métodos se puede encontrar en la siguiente referencia [7].

Este apartado, se basa en la *Encuesta de inserción laboral de titulados universitarios* del INE. Estos datos contienen información de los titulados universitarios de 1er y 2º ciclo y graduados del curso 2009-2010 en todas las universidades del territorio nacional. El período de información es el comprendido entre la finalización de los estudios y el momento de la entrevista. Estos contienen información de 30000 titulados aproximadamente, conteniendo 181 variables.

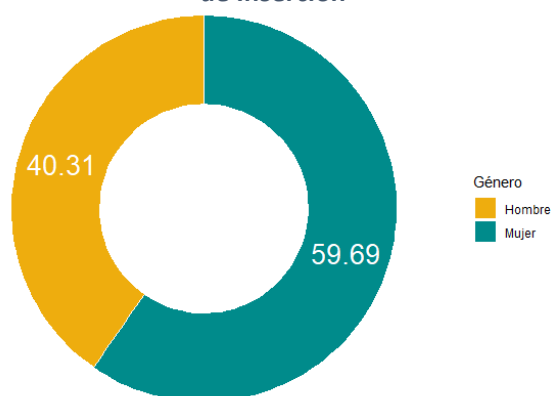
Se han considerado las reglas de asociación una buena manera de abordar este tipo de datos, puesto que los datos provenientes de encuestas siempre tienen la limitación del flujo concreto

que siguen. Según el caso de cada individuo, se sigue un flujo concreto de preguntas, lo que provoca que muchas variables tengan casos con “No aplicables” y eso, a veces, dificulta según que tipo de análisis. Es por esto, que se cree que las reglas de asociación pueden ser favorables en el sentido de que, por un lado, se generen reglas que recojan estos flujos “no válidos” de los casos donde no se ha pasado por unas variables concretas y, obviando estas reglas, se puedan ver patrones de comportamiento con las demás, es decir, que el flujo natural de la encuesta se agrupe con esta metodología.

Para evitar que el hecho de tener mayor presencia de mujeres en los datos influya en los resultados, se ajustarán reglas con un *lift* mayor a 1, esperando encontrar aquellas que tengan un comportamiento diferente al esperado en caso de independencia y no tanto aquellas que se den mucho en los datos, en resumen, aquellas donde se encuentre dependencia entre las categorías implicadas.

Finalmente, comentar que en este capítulo se encontraría el mismo hecho comentado sobre la

Figura I-2. Proporción de géneros en los datos de Inserción



Fuente: Elaboración propia

prueben.

EADA. Puesto que en las universidades hay una mayor presencia de mujeres, estas son mayoría en la muestra y, por lo tanto, en los gráficos presentados para comparar la presencia de hombres y mujeres, siempre se encontrará mayoría de mujeres, no por diferenciaciones de género, sino por un mayor nombre absoluto de ellas en los datos. De nuevo, destacar que los gráficos son un punto de apoyo, pero siempre que se determine la existencia de diferencias, se hará a través de métodos estadísticos que lo

3. Análisis de concienciación

En este último capítulo se han empleado técnicas no tanto de tipo estadístico, más bien del ámbito informático, entrando muy levemente en aspectos de Big Data. Con la finalidad de obtener datos de Twitter, en primera instancia se creó un perfil de desarrolladora de Twitter, para tener acceso a la API. Posteriormente se llevó a cabo la conexión entre RStudio y este acceso a los datos, para poder extraer los tweets y poder trabajar con ellos en el entorno de R. Todo esto gracias a las siguiente referencias [8] [9] [10].

También se ha realizado un análisis de series temporales con los datos de Google, extraídos de Google Trends. Con la finalidad de conocer la existencia de tendencia o estacionalidad si la había, se ha empleado, en cuanto a recursos gráficos, descomposición de la serie con esquema aditivo con la base teórica en [11]. El test empleado para la comprobación estadística de la tendencia ha sido el test de Spearman [12].

Recursos informáticos

Finalmente, de manera general, se hará mención a los recursos informáticos. El programa principal del trabajo ha sido R, la herramienta más conocida y usada durante el grado y más enfocada a la estadística por el momento. Si bien es cierto que esta se conocía por su gran uso durante toda la carrera, se ha querido profundizar en su uso, tratando de exprimir sus funcionalidades para unos mejores resultados. En cuanto a las representaciones gráficas de todo el trabajo, se han llevado a cabo mediante la librería ggplot2(), explorada en profundidad con la intención de sacarle el máximo partido y obtener una buena visualización de los datos, que proporcionen un buen entendimiento de los resultados.

Junto con ggplot2, las principales librerías para la realización de todos los procesos llevados a cabo han sido dplyr, FactoMineR, twitterR, tidyr, wordcloud, arules, entre otras muchas, que han facilitado el trabajo realizado, juntamente con los recursos públicos que se encuentran en internet que son de mucha utilidad.

Documentación

En primer lugar, se llevó a cabo un proceso de documentación en estudios estadísticos realizados sobre diferencias de género y otros que pudieran servir de inspiración y de punto de partida para este trabajo.

Los principales artículos son:

- ✓ Brechas salariales de género en España [1]
- ✓ Ser estudiant avui – Via Universitaria [14]
- ✓ The Determinants of Motherhood and Work Status [15]
- ✓ Gender Gaps in the Spanish Labor Market [16]
- ✓ Understanding International Differences in the Gender Pay Gap [17]

Este primer trabajo estudia la evolución de las brechas salariales de género ajustadas mediante la *Encuesta de Estructura Salarial*.

La brecha salarial española se ha reducido un 33% desde 2002, aunque se siguen detectando dinámicas preocupantes y algunas de las que más se relacionan con este trabajo son: los hombres con educación secundaria ganan un 14% más y los universitarios un 12% más que las mujeres; un hombre con más de 7 años de antigüedad en la empresa gana un 16,4% más que una mujer con la misma antigüedad; los hombres gana un 14,3% más que las mujeres en las empresas grandes y un 9,4% más en las empresas pequeñas; en cuanto mayor es el porcentaje de mujeres (tanto en ocupaciones como en sectores de actividad) menor es la brecha ajustada.

Sobre otro ámbito de la vida profesional de una persona nos informa el artículo *Ser estudiant avui*, que analiza el acceso, las condiciones de aprendizaje y las expectativas de los estudiantes universitarios. Este se centró en analizar un campo muy concreto, que también aparece en este trabajo, como son los estudiantes universitarios. De este estudio se extraen los siguientes resultados: las mujeres son mayoría en las universidades (62%) pero su presencia es minoritaria en Ingenierías (32%) o titulaciones mixtas (44.6%); las mujeres muestran un nivel de dedicación en el estudio (más de 40h semanales) mayor (40%) frente a los hombres (34%).

Cambiando de enfoque, se presenta el artículo "The Determinants of Motherhood and Work Status". Este dice que En las últimas décadas la participación de las mujeres en el mercado de trabajo ha registrado una tendencia alcista, a la vez que la fertilidad se ha reducido considerablemente. También ha incrementado la participación de las mujeres en cursos universitarios. Para conseguir esto, las mujeres han retrasado la edad del matrimonio, a la vez que la edad en la que tienen hijos.

Sin embargo, desde medianos de 1980, muchos estudios han descubierto que está correlación anteriormente negativa se ha vuelto positiva. Este cambio se debe a los cambios en el estado del bienestar, y las políticas microeconómicas que mediante ayudas a la conciliación entre el hecho de ser madre y el trabajo (Facilitan la flexibilidad laboral, etc..), mediante la reducción del coste de oportunidad de tener hijos, facilitan la posibilidad de ser madres y a la vez participar en el mercado de trabajo.

Volviendo a la brecha salarial, se entra en dos artículos más, uno más genérico del ámbito español, que proporcionará un mejor entendimiento del entorno y el siguiente sobre brecha salarial internacionalmente.

Del primero se extrae que España es uno de los países europeos con un modelo familiar más tradicional. Aproximadamente el 30% de los hogares en España siguen el modelo tradicional de una pareja con hijos. Otros países europeos como por ejemplo Alemania bajan este número hasta el 20% aproximadamente.

En cuanto a la educación, se puede ver que en edades ya muy avanzadas (a partir de 55 años los hombres presentan un nivel educativo más alto que las mujeres en España, sin embargo, en las generaciones inferiores a 55 años son las mujeres las que tienen un nivel de estudios mayor incrementando este gap en las generaciones más jóvenes (menores de 30 años) a 13 puntos porcentuales de diferencia entre mujeres graduadas en estudios universitarios respecto los hombres.

A la vez también existe una diferencia entre hombres y mujeres muy clara en el campo de especialización escogido. Las mujeres constituyen el 70% de graduadas en áreas como la salud, la educación y los servicios sociales. En cambio, menos del 30% de graduados en ingenierías o informática son mujeres.

Finalmente, entrando en el ámbito internacional de la brecha salarial se obtiene lo siguiente. En las últimas décadas este *gap* ha disminuido significativamente, sin embargo, todavía existe una diferencia entre los salarios de hombres y mujeres, y también existen diferencias en el *gap* respecto diferentes países. Los economistas tradicionalmente han atribuido éste *gap* a aspectos específicos del género o discriminación por parte del empleador. Sin embargo, comenzando en 1991 con un estudio de Juhn, Murphy and Pierce, los economistas han reconocido que la estructura salarial en sí, o el hecho de que los precios del trabajo van pegados de las habilidades personales de cada trabajador pueden tener un mayor impacto en el *gap* entre diferentes subgrupos, como el del género. Por ejemplo, como de media las mujeres tienen menos experiencia que los hombres en el mercado laboral, y trabajan en

sectores diferentes, siendo todas las otras variables iguales, existirá una diferencia en los salarios entre ambos grupos.

Más soporte de que el gap entre los salarios entre géneros está creado principalmente por la estructura salarial en sí, lo encontró Rowthorn, basándose en tabulaciones de datos agregados de los salarios de 17 países entre 1973 y 1985, dónde la diferencia entre géneros era más pequeña en aquellos países con una centralización mayor de la estructura salarial.

Los resultados muestran que estructuras salariales igualitarias favorecen una brecha salarial menor. También que las diferentes instituciones y políticas utilizadas en diferentes países son un factor de mucha importancia a la hora de determinar estructuras salariales igualitarias, y por lo tanto a la hora de determinar la brecha salarial de un país.

Cuerpo del trabajo

A continuación, se presenta el contenido práctico del trabajo, donde se irán explicando los análisis llevados a cabo y los resultados obtenidos en los diferentes apartados. Este se divide en tres capítulos: formación, inserción laboral y redes sociales.

En el primer bloque se analizarán variables de educación y formación de los individuos españoles, tanto la educación regalada de estos, como la formación continuada durante la carrera profesional, así como el aprendizaje informal de las personas.

En el siguiente capítulo se abordará el aspecto de la inserción laboral de los titulados universitarios, estudiando aspectos sobre el proceso de cambio entre la educación y la vida profesional.

Finalmente, en el bloque de redes sociales, se analizará, por una parte, la interacción en Twitter en torno a diversos *Hashatgs*² relacionados con la temática de estudio; y, por otro lado, el interés de búsquedas en Google de diversos términos también relacionados.

Capítulo I. ¿Brecha educacional?

En este primer apartado, se analizarán las características educativas de los individuos españoles.

Se busca analizar si hay diferencias en el ámbito educativo entre hombres y mujeres. Observar si estos participan de manera diferente en la educación, si estudian cosas diferentes o, en general, si hay puntos donde se produzca una *brecha estudiantil* que pudiera contextualizar y acompañar a la existente *brecha salarial*.

Para el mejor entendimiento se estudiarán dos grandes apartados sobre la formación de una persona, como son la educación formal y reglada, y la no formal.

Estos dos grandes temas se definen con la intención de separar entre estudios reglados y oficiales y formación continua personal, no tanto en la época estudiantil de una persona, sino a lo largo de su carrera profesional o posterior a sus estudios oficiales.

² Etiqueta, es una cadena de caracteres precedidos por una almohadilla con el fin de que se identifique la temática en cuestión. Se usa en diversas redes sociales

1.1. Educación formal

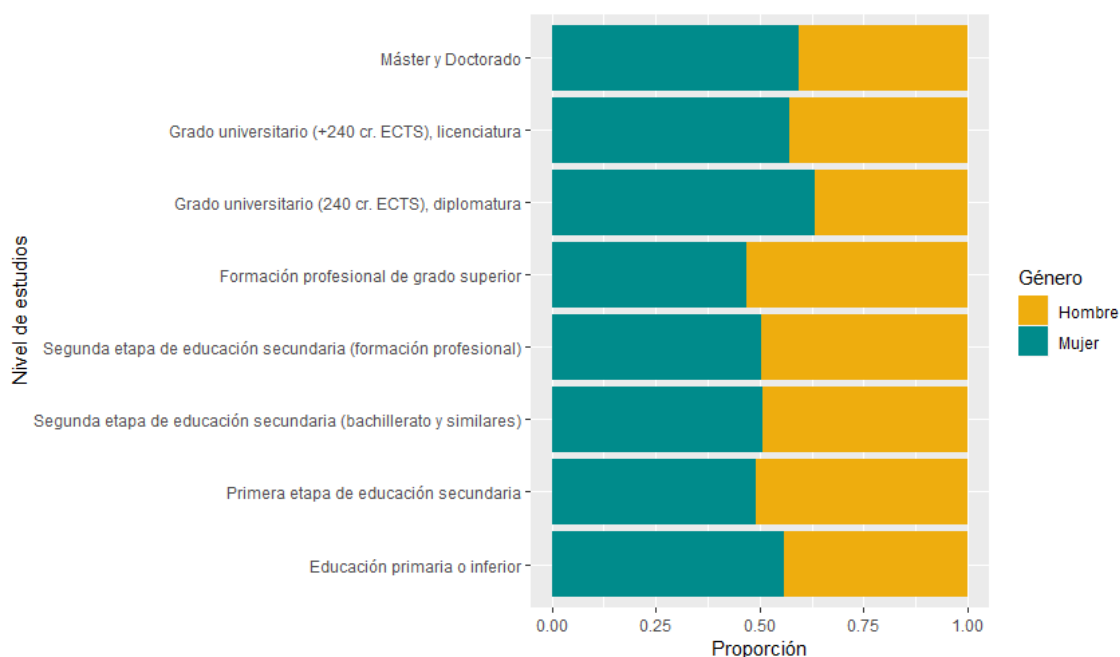
En este primer apartado se analizarán las variables referentes a la educación formal³ de un individuo, comparando siempre ente géneros.

A grandes rasgos, las variables que se han analizado hacen referencia al nivel de estudios y al contenido de estos, también tratan sobre si actualmente la persona en cuestión es estudiante, las razones para serlo y características de estos estudios.

Debido al gran contenido de la encuesta, se ha hecho un filtraje previo de un primer grupo de variables consideradas más representativas para el objeto de estudio; posteriormente, las que pueden aportar conclusiones o ideas más relevantes para la detección, o no, de diferencias entre géneros, serán las presentadas en el trabajo.

Para el máximo nivel de estudios diferenciando entre mujeres y hombres no se observan grandes diferencias, aunque sí que se puede ver una mayor presencia de mujeres en grados universitarios y másteres, así como mayor presencia de hombres en estudios de más bajo nivel como formación profesional o la educación secundaria. Estos comportamientos reafirmarían los observados en otros estudios, donde se percibe que hay más mujeres en las universidades.

Figura I-1. Presencia de géneros en el máximo nivel de estudios



Fuente: Elaboración propia

³ Se considera educación formal aquellas actividades realizadas dentro del sistema educativo con las que se obtiene un título oficial

Para determinar si hay diferencias significativas entre sexos en cuanto al nivel de estudios se realiza un test χ^2 del cual se obtienen el siguiente resultado. Se observa como el p-valor es

Ilustración I-1. Resultado obtenido del test de chi-cuadrado (R)

```
Pearson's Chi-squared test
data:  tabla
X-squared = 214.84, df = 7, p-value < 2.2e-16
```

menor a 0.05, por lo tanto, el test resulta significativo, lo que indica que hay incidencia del género en el máximo nivel de estudios de una

persona. Para conocer que categorías son más determinantes en esta diferenciación se observa la tabla de residuos.

Cada columna se corresponde con una categoría, donde el 1 es “Educación primaria o inferior” y 8 “Máster y doctorado”.

Tabla I-1. Tabla de residuos obtenida del test chi-cuadrado

	1	2	3	4	5	6	7	8
Hombre	-1.83	4.03	1.21	1.03	3.46	-7.54	-3.18	-3.31
Mujer	1.75	-3.86	-1.16	-0.99	-3.31	7.23	3.05	3.17

Fuente: Elaboración propia

El mayor residuo se obtiene, con diferencia, para la categoría de *Grados universitarios de 240 créditos ECTS o diplomados universitarios*. También con mayores residuos respecto al resto, aunque alejadas de este primero, están las categorías *primera etapa de educación secundaria, formación profesional de grado superior, grado universitario de más de 240 créditos ECTS o licenciatura y Máster y doctorado*.

Si se analiza, por otro lado, las desviaciones respecto a los valores esperados en caso de independencia se obtiene la siguiente tabla.

Tabla I-2. Tabla de desviaciones obtenida del test chi-cuadrado

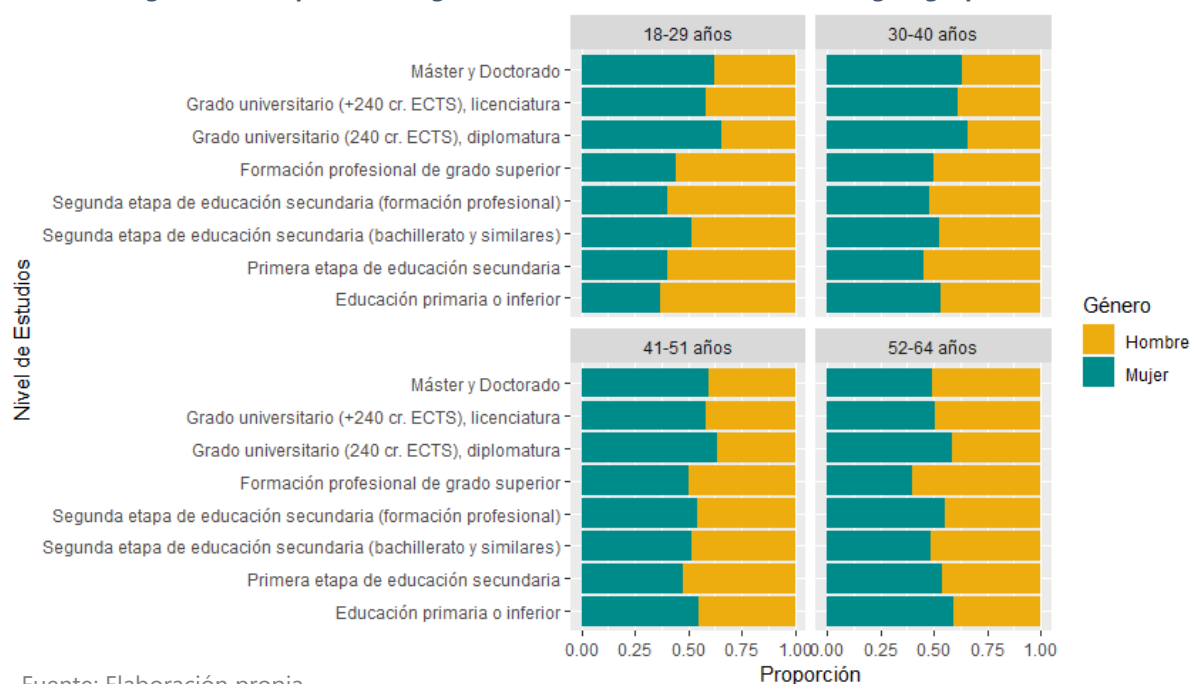
	1	2	3	4	5	6	7	8
Hombre	-2.59	6.99	1.86	1.50	5.02	-10.96	-4.61	-4.68
Mujer	2.59	-6.99	-1.86	-1.50	-5.02	10.96	4.61	4.68

Fuente: Elaboración propia

Con esta otra, se reafirman las ideas extraídas mediante los residuos. Con estas dos tablas se observa que, en las categorías más “extremas” como son los estudios universitarios o la de educación primaria es donde se encuentra mayor presencia de mujeres, mientras que en las categorías intermedias de estudios secundarios o formación profesional están más masculinizadas.

A continuación, añadiendo el cruce según la variable de edad agrupada, se acabará de observar este fenómeno.

Figura I-2. Proporción de géneros en cada nivel de estudios según grupo de edad



Fuente: Elaboración propia

Diferenciando por grupos de edades, se puede observar este cambio de tendencia, donde a mayor edad mayor presencia de mujeres en niveles de estudios más bajos y a medida que esta se reduce, va percibiéndose una proporción más elevada de mujeres en estudios superiores.

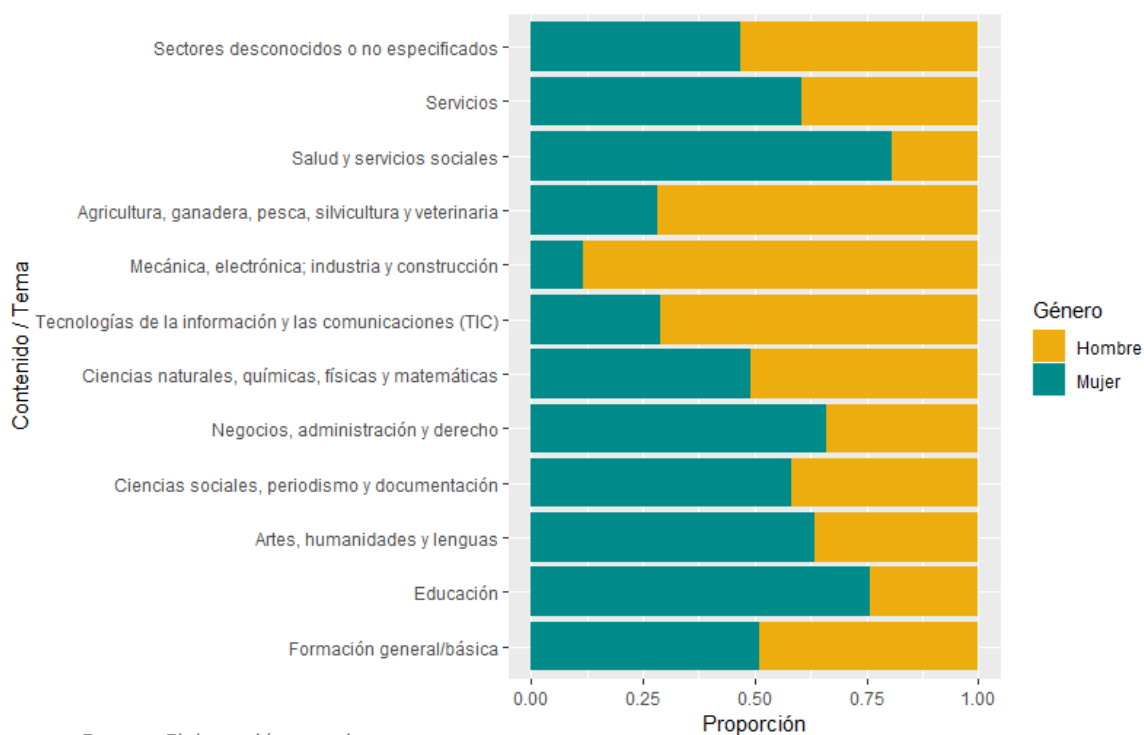
De aquí se concluiría que las mujeres obtienen o bien un nivel bajo de estudios, donde se centran las mujeres de mayor edad, o bien un alto nivel de estudios.

Se podría extraer la siguiente idea de que, al históricamente no tener acceso las mujeres a la educación o solamente a la educación más básica, estas se concentraban en niveles como analfabetismo o estudios primarios. A medida que el género femenino ha ido ganando en derechos hacia la igualdad, su presencia ha ido incrementando en los niveles más altos de los estudios, posicionándose actualmente como la mayoría de los y las estudiantes de grados y másteres. También cabe destacar que, a medida que aumenta la edad, la presencia de mujeres en estudios intermedios como formación profesional o bachillerato es cada vez menor, que se podría relacionar con las profesiones más manuales o físicas, muy masculinizadas, y que en la mayoría de los casos no requieren de estudios de muy alto nivel.

Si de esta formación se observa el contenido o tema según el sexo, se observan los claros sectores masculinizados y feminizados, tanto gráficamente como mediante el contraste de

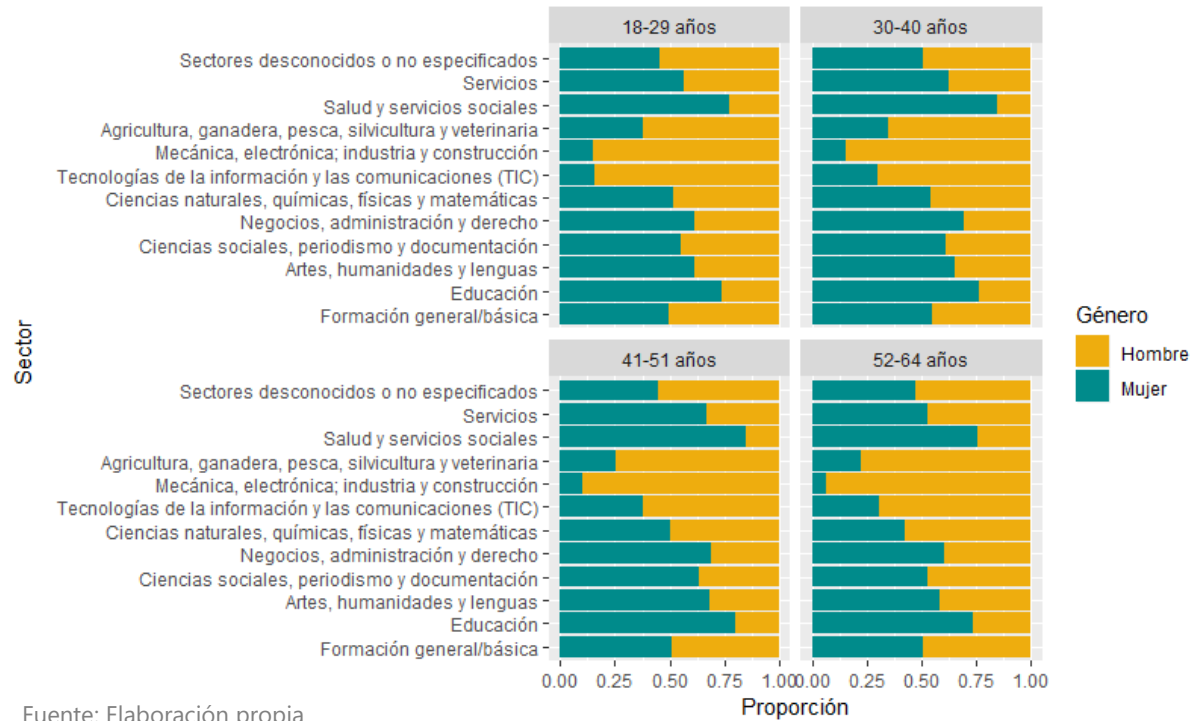
hipótesis. Las temáticas que más inciden en estas diferencias son la mecánica, electrónica, industria y construcción, sectores altamente masculinizados; por otro lado, la salud y los servicios sociales, muy feminizados. También las áreas de educación, artes, humanidades y lenguas y los negocios y la administración, con más presencia de mujeres y las TIC con mayor proporción de hombres.

Figura I-3. Presencia de géneros en cada temática de estudio (base: 13569)



Esta variable también resulta interesante observarla por edades, con la intención de ver si estas diferenciaciones se reducen en las nuevas generaciones o todavía persisten.

Figura I-4. Presencia de géneros en cada temática de estudio según grupo de edad (base: 13569)

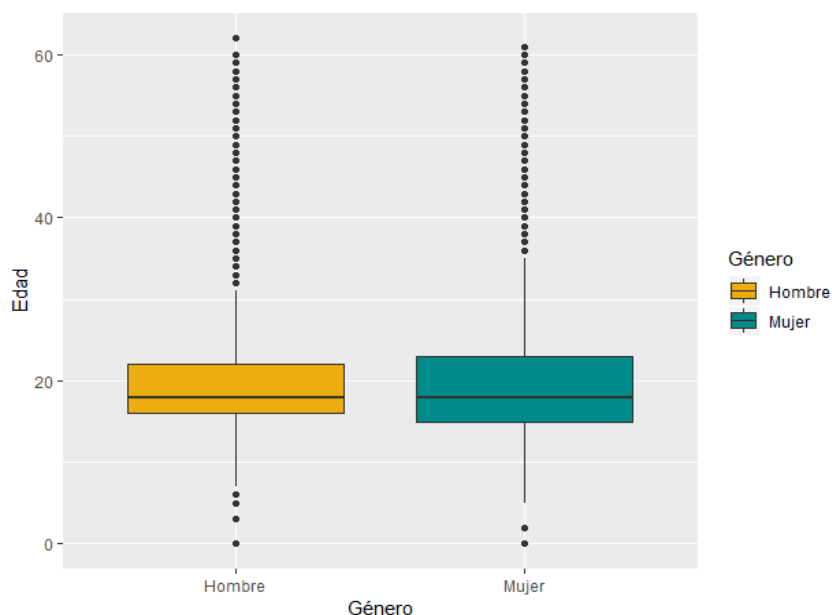


Fuente: Elaboración propia

Se percibe que la proporción de mujeres y hombres en los contenidos es bastante similar para todos los grupos. Con esto se podría decir que las diferencias marcadas entre áreas temáticas de hombres y de mujeres existe tanto en las personas mayores y de edad adulta que estudiaron hace tiempo como para las nuevas generaciones, por lo tanto, este sería un aspecto con un largo camino por recorrer hacia la disminución de diferencias entre géneros.

A cotinuación, se observa la distribución de la edad de finalización de los estudios.

Figura I-5. Distribución de edades según género (base: 20871)



Fuente: Elaboración propia

Visiblemente se percibe una mayor variabilidad en la edad de finalización de las mujeres, aunque la mediana de edad se encuentre en los 19 años para ambos géneros.

Para determinar si hay diferencias entre hombres y mujeres, se realiza un test t-student. A

Ilustración I-2. Resultado obtenido del test t-student (R)

```
welch Two Sample t-test
data: HATAGE by SEX
t = -3.2358, df = 20839, p-value = 0.001215
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4759169 -0.1168477
sample estimates:
mean in group Hombre mean in group Mujer
      19.67620          19.97258
```

continuación, se presenta el resultado obtenido, donde se puede ver que el test resulta significativo, indicando que la edad de finalización de sus estudios de las mujeres es mayor a la de los

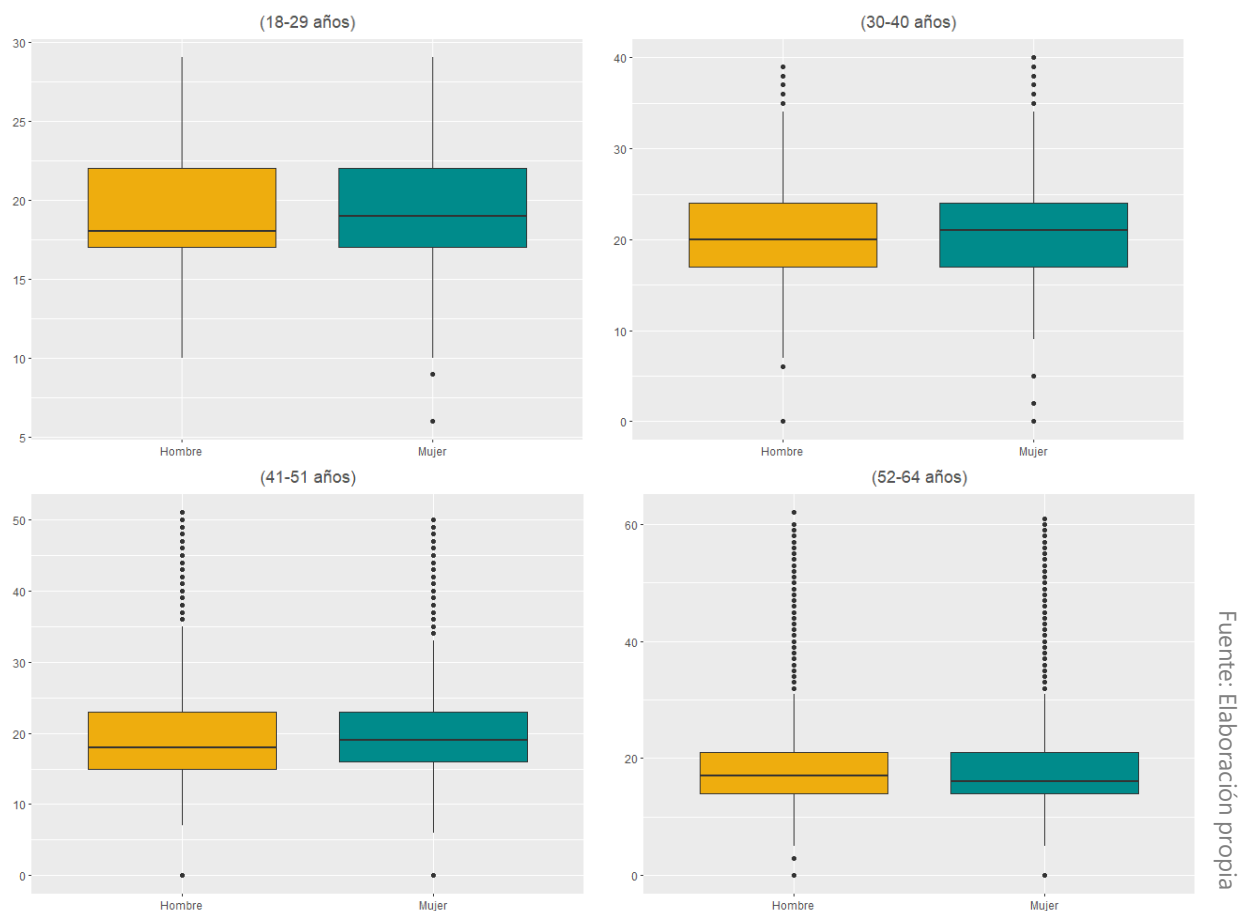
hombres.

Con esto, se determina que ellas se demoran más en acabar sus estudios, hecho que se podría relacionar con la mayor presencia de mujeres en niveles de estudios superiores. Al llegar las mujeres a obtener un mayor nivel de estudios, la media de edad de finalización global de las mujeres resulta mayor que la de los hombres.

Por otro lado, esta variable también tiene mucho sentido mirarla por grupo de edad, ya que, como se viene observando, los comportamientos de mujeres y hombres varían según la edad de estos y las diferencias entre ellos, también.

Si esto mismo se replica para cada grupo de edad definido, se puede ver alguna diferencia en cuanto a la mediana de edad, principalmente en los grupos de menor edad, habiendo una diferencia de 1 año en los tres rangos de edad menores.

Figura I-6. Distribución de edades por género según grupo de edad (base: 20871)

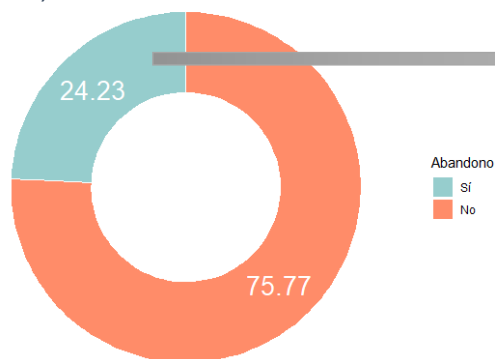


Realizando el test para cada grupo de edad, este resulta significativo para los grupos de edad de 18-29 años y 41-51 años, mientras que no lo es para los otros dos. Estos resultados no son vinculantes con una explicación social clara, con lo que son sujetos a un posible estudio más profundo, para entender el por qué estas diferenciaciones entre grupos de edad. Puesto que el objeto de estudio son las diferencias entre género y no tanto entre grupos de edad, este fenómeno no se analizará más allá.

Otro aspecto de estudio es el abandono de alguna enseñanza oficial, posterior al nivel máximo de estudio. Un 25% afirma haber abandonado alguna otra enseñanza diferente a la oficial y, de estos, un 52.41% serían hombres.

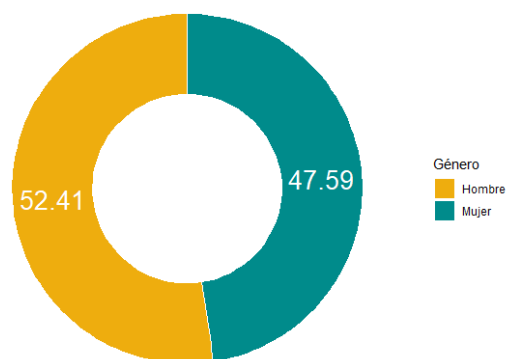
Esto indicaría que son ellos los que más han abandonado alguna educación posterior, con casi 5 puntos de diferencia.

Figura I-7. Abandono enseñanza oficial
(base: 8206)



Fuente: Elaboración propia

Figura I-8. Géneros que abandonaron enseñanza

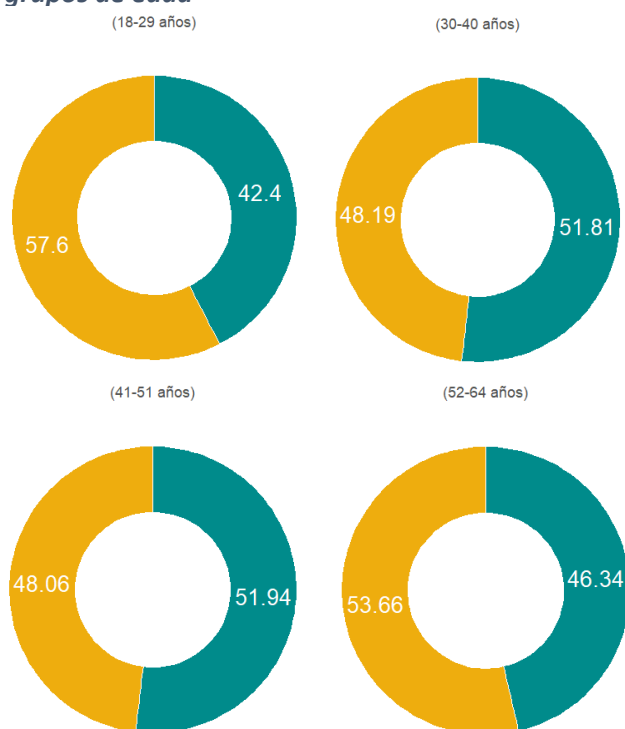


Fuente: Elaboración propia

Mediante el test, el cual resulta significativo, se confirman diferencias entre hombres y mujeres en torno al abandono de actividades educativas.

Una vez más, para entender bien este suceso, se observa la variable para los diferentes grupos

Figura I-9. Géneros que abandonan enseñanza por grupos de edad



Fuente: Elaboración propia

abandono, también en relación con los claros sectores masculinizados, donde en muchos casos no sería necesario un nivel de estudios superior.

de edad. Se observa como en los grupos de edad más extremos, son los hombres quienes representan mayoría al abandonar una enseñanza posterior; mientras que, en los dos grupos de edad centrales, ellas representan una leve mayoría.

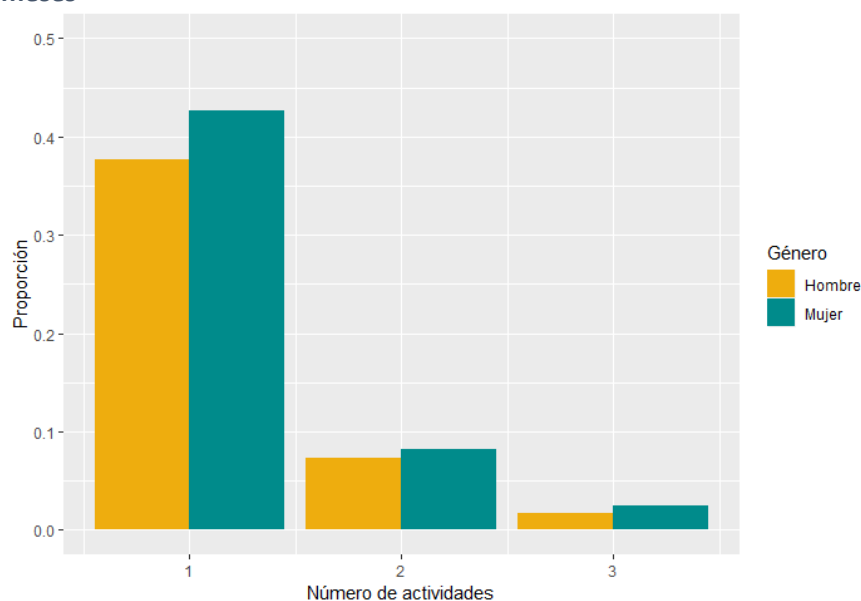
Ese suceso se podría relacionar con lo observado en torno al nivel de estudios. En las edades extremas se producen las diferencias más claras; para el grupo de mayor edad, las mujeres se quedaban en un nivel bajo de estudios siendo mayoría, mientras que en el grupo más joven, ellas son mayoría en los niveles de estudio superiores. Consecuentemente, el sexo opuesto, tendría una mayor tasa de

Una vez introducidas las variables más generales en cuanto a la educación de una persona, se hará un análisis más detallado sobre las personas que han cursado estudios formales en los últimos doce meses.

En estas variables no se entrará tanto al detalle de observar las diferencias de género en los grupos de edades, porque no se considera que en este caso sea tan decisiva la edad de una persona. En estas variables no participa el factor temporal de la etapa en la que se estudió, puesto que el objetivo de estas variables son los estudiantes actuales.

En los últimos 12 meses un 15% de la población ha participado o está participando en una actividad educativa oficial. Si esta subpoblación se diferencia por géneros se obtiene el siguiente gráfico.

Figura I-10. Presencia de géneros en actividades en los últimos 12 meses



En este se ven representados los individuos que han dicho estar cursando una, dos o tres actividades actualmente o en los últimos doce meses. Se observa una leve mayor presencia de mujeres, acentuándose en los individuos que cursan o han cursado una actividad.

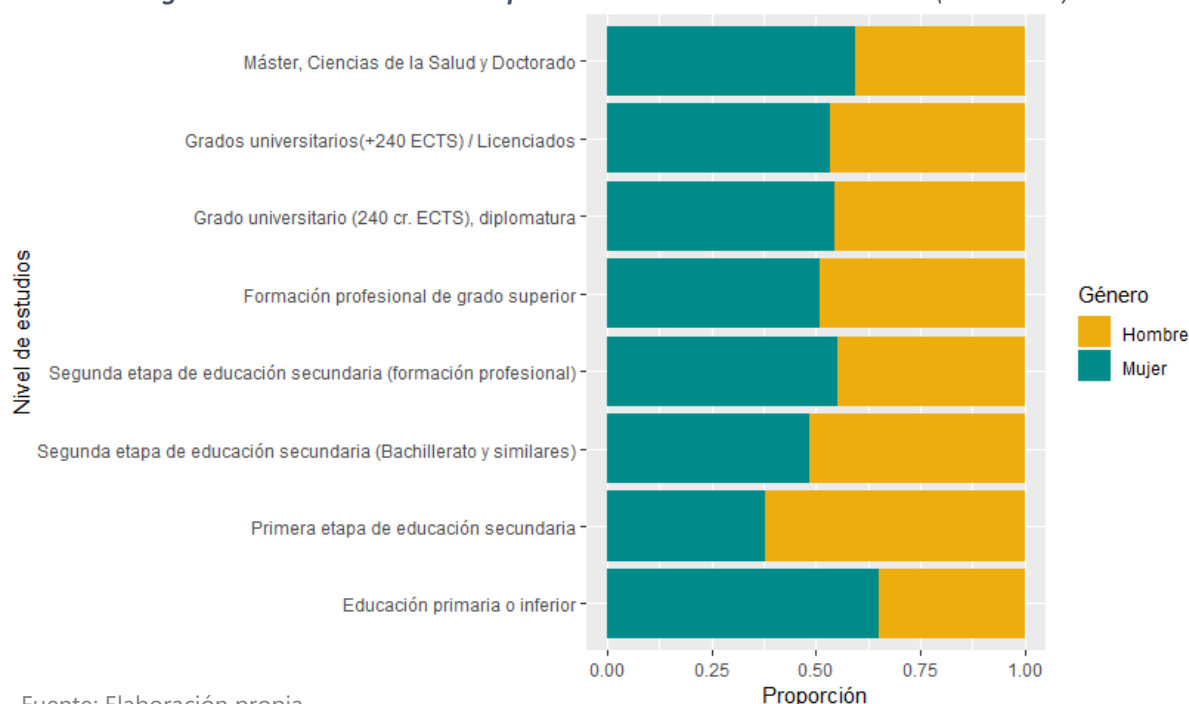
Fuente: Elaboración propia

Sin embargo, al realizar el contraste de hipótesis, este no resulta significativo. No existen diferencias de géneros en el número de actividades educativas realizadas entre los estudiantes actuales, ambos géneros se comportan igual en torno al número de actividades formativas que realizan.

Cabe destacar que los casos que estudian 1 actividad son 2761, que estudian 2 son 536 y los que estudian 3 son 143, por lo tanto, estos datos deben ser interpretados con cautela, por los pocos individuos que conforman algunas categorías.

Anteriormente, se han observado diferencias tanto en el nivel de estudios como en el contenido de estos, y estas diferencias se percibían tanto a nivel global, como observando los diferentes grupos de edades. A continuación, se estudiarán estas mismas variables en los estudiantes actuales.

Figura I-11. Nivel de estudios que cursan los actuales estudiantes (base: 3440)



Con esta primera variable, que hace referencia al nivel que están cursando los estudiantes actuales, se pueden observar diversos fenómenos que ya se han ido comentando.

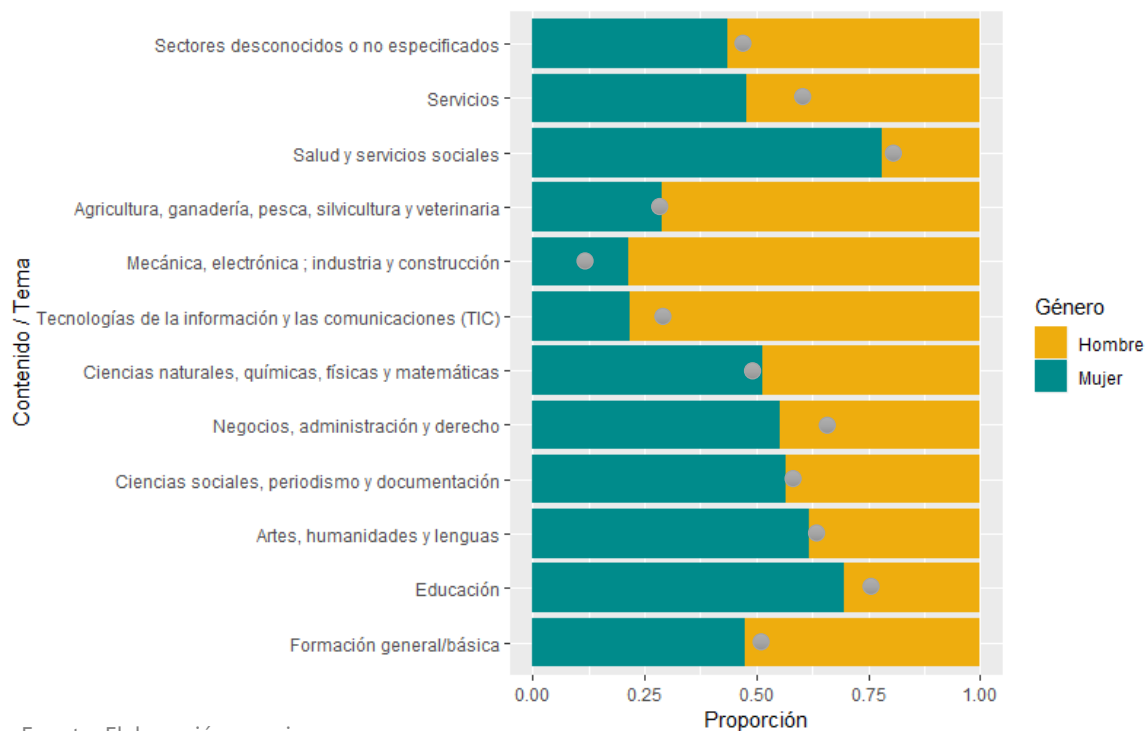
El primero sería que, debido a que las mujeres de mayor edad representaban mayoría en los niveles más bajos de estudios, son estas las que ahora representan mayoría de estudiantes actuales en este mismo nivel. Es decir, como en su momento no estudiaron o estudiaron muy poco, son ellas las que tienen más peso a la hora de llegar a este mínimo de educación ya de mayores.

Por otro lado, se observa esta leve tendencia de que las mujeres se concentran en los niveles más altos de estudio, mientras ellos tienen más presencia en niveles intermedios.

De nuevo, resulta significativo el hecho de que hay diferencias entre hombres y mujeres en lo que respecta al nivel de estudios que estos adquieren, o están adquiriendo. Esto nos indica que, no es solo una problemática histórica, sino que todavía ocurre entre los actuales estudiantes.

También se observan los contenidos y temáticas de los estudiantes actuales.

Figura I-12. Temática de los estudios de los actuales estudiantes (base: 3245)

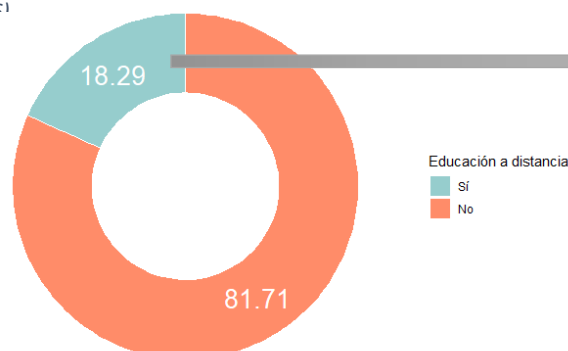


A simple vista, se siguen viendo esas claras temáticas feminizadas y masculinizadas, ya comentadas anteriormente y estas diferencias también resultan significativas entre los actuales estudiantes. Los círculos grises representan donde se situaba la “separación” entre hombres y mujeres en el gráfico general. Se puede ver como las proporciones se han mantenido igual o incluso han empeorado, generándose una mayor desproporción entre hombres y mujeres; solo en la temática de *mecánica, electrónica, industria y construcción* se detecta un leve aumento en el porcentaje de mujeres que forman parte de estudios relacionados con dicha temática.

Con esto, se podría decir que sigue habiendo una gran brecha temática entre hombres y mujeres.

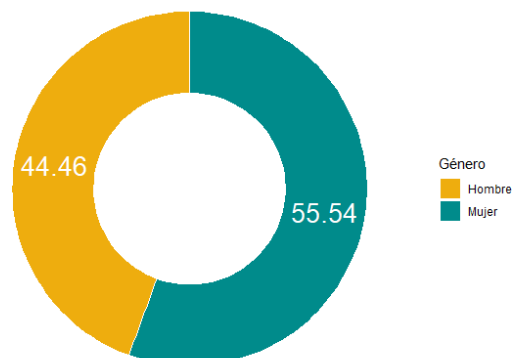
Si sobre esta educación, se estudia si esta era principalmente a distancia o no, se observa cómo el 18.29% de los estudios se realizan a distancia, de los cuales un 44.46% lo conforman hombres y el 55.54% restante mujeres.

Figura I-14. Cursando educación a distancia (base: 24061)



Fuente: Elaboración propia

Figura I-13. Educación a distancia según género (base: 24061)

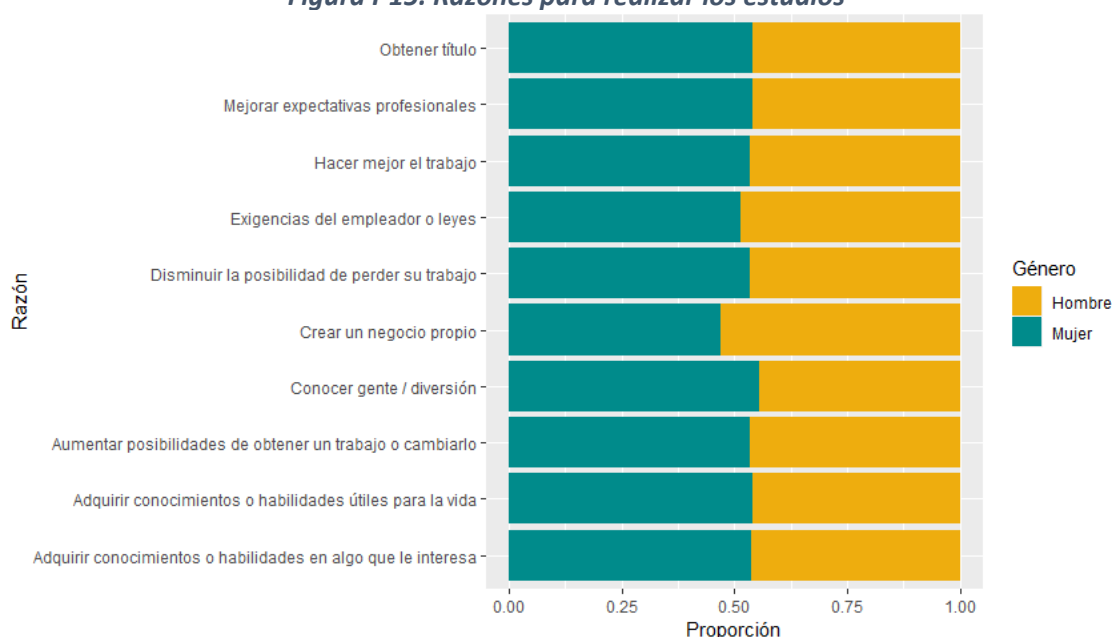


Fuente: Elaboración propia

Aunque visualmente se percibe una diferencia de aproximadamente 11 puntos entre los géneros, estas diferencias no resultan significativas. Por lo tanto, en torno a la actitud de participar en estudios a distancia no habría diferencias de género.

Sobre las razones para cursar estos estudios, no se ven grandes diferenciaciones a priori, pero si habría algunas razones disputantes. Estas serían *disminuir la posibilidad de perder el trabajo* y *conocer gente / diversión* con una mayor proporción de mujeres, mientras que la razón donde se encuentran más hombres es en *crear un propio negocio*.

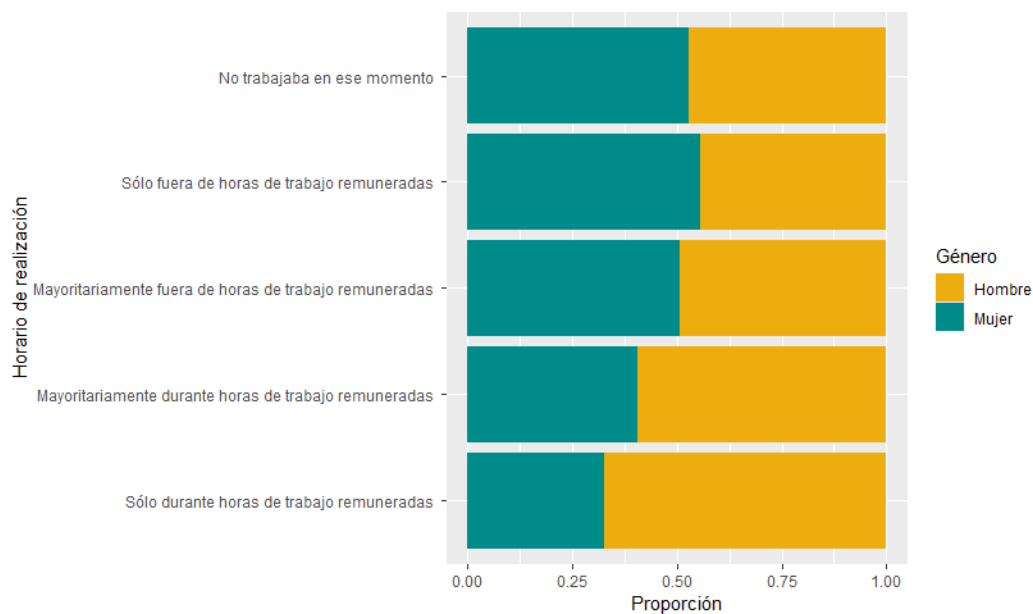
Figura I-15. Razones para realizar los estudios



Fuente: Elaboración propia

Sobre estas actividades educativas, se analiza el hecho de haberlas realizado en horas de trabajo remuneradas y, de nuevo, diferenciando entre géneros se observa lo siguiente.

Figura I-16. Actividad realizada en horas de trabajo remuneradas (base: 3297)



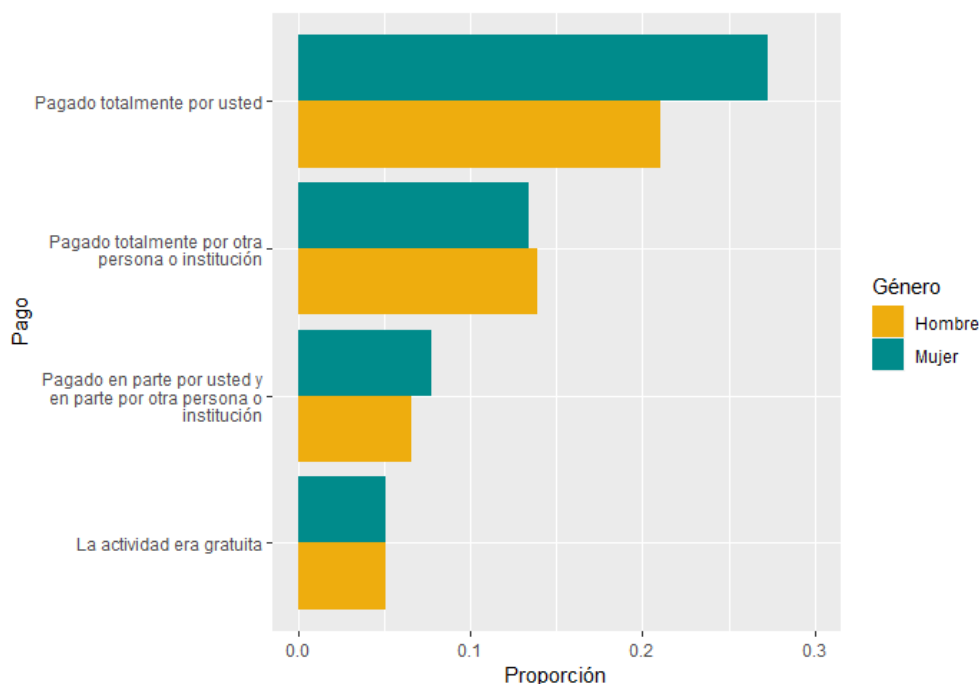
Fuente: Elaboración propia

A simple vista se percibe una relación gradual entre el género y la dedicación en el trabajo o no. Entre los que sólo se dedican a la actividad durante horas de trabajo remuneradas, se encuentra una amplia mayoría masculina, la cual se va reduciendo a medida que las horas en dedicadas en el trabajo son menores. En la categoría opuesta, que es solo dedicarse a la actividad fuera de horas de trabajo remuneradas, se encuentra una proporción más igualitaria

Esta representación ya apunta a una posible diferenciación entre géneros a la hora de estudiar en el puesto de trabajo, que queda confirmada con la realización del test de Pearson, el cual resulta significativo e indica una relación entre el género de una persona y las horas que esta dedica a su actividad educativa en el trabajo.

En cuanto al pago de la actividad educativa, se puede ver cómo un 27%, aproximadamente, de las mujeres se pagan la totalidad de los estudios, mientras que para los hombres esta modalidad de pago representa el 21% aproximadamente. También se puede percibir una mayor presencia, aunque leve, de mujeres que combinan el pago propio y por otra persona y, en cambio, una mayor presencia masculina cuando el pago es de manera totalitaria por otra persona o institución.

Figura I-17. Pago de los gastos de la actividad (base: 3378)

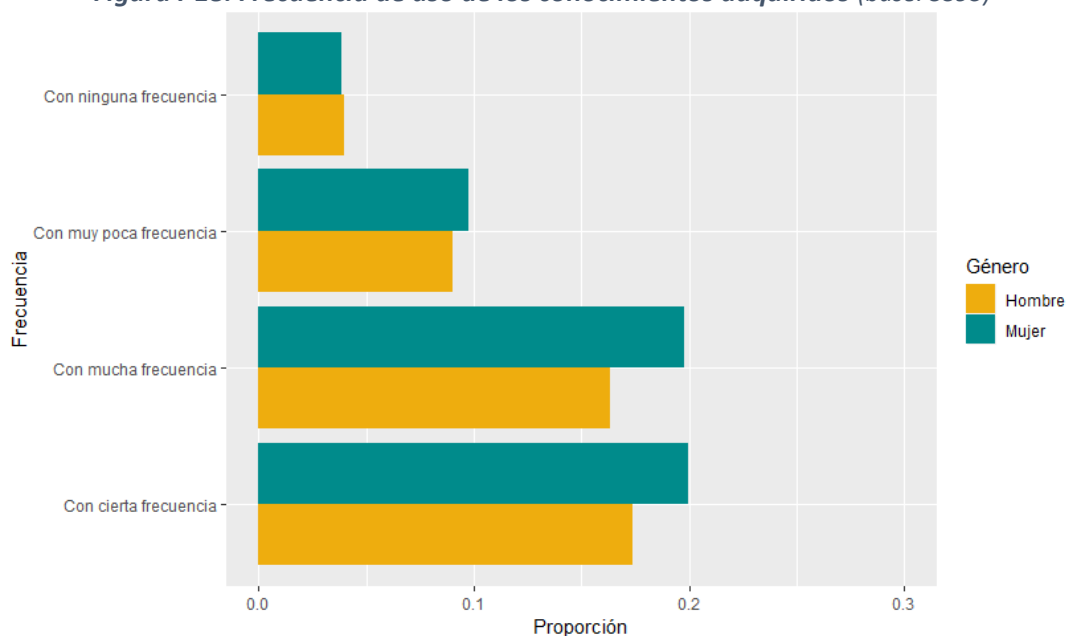


Fuente: Elaboración propia

Además de percibirse estas diferencias a la hora de representarlas, también resultan significativas, con lo que se podría concluir que hay diferencias de género en la forma de pago de las actividades educativas entre los estudiantes actuales. Siendo las mujeres quienes más se pagan los estudios de manera íntegra ellas mismas o combinando el pago con otra persona/institución; mientras que son los hombres quienes más disfrutan de actividades educativas pagadas o gratuitas.

En referencia a la aplicación de los conocimientos adquiridos en la actividad educativa, se les pregunta a los estudiantes sobre la frecuencia de uso de los conocimientos y habilidades adquiridos.

Figura I-18. Frecuencia de uso de los conocimientos adquiridos (base: 3395)

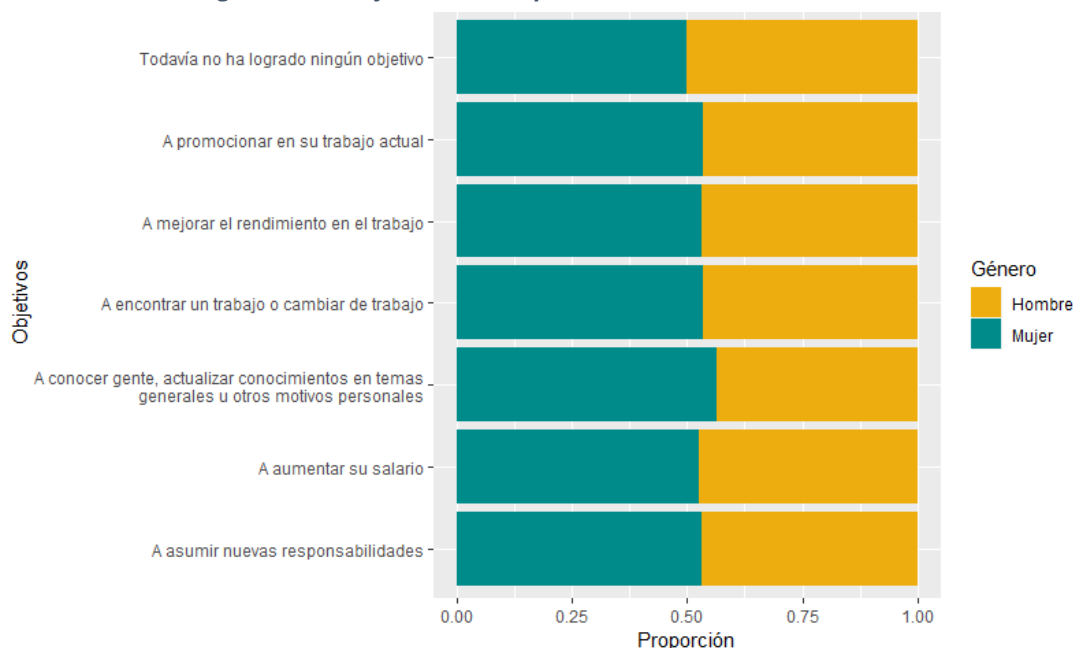


Fuente: Elaboración propia

Gráficamente puede verse una mayor presencia de mujeres en las categorías, pero no habría diferencias significativas. Hombres y mujeres consideran usar los conocimientos adquiridos con la misma frecuencia.

Finalmente, en cuanto a los objetivos a los que ha contribuido la actividad no se observan proporciones muy diferenciadas. Solo se podría destacar una leve mayor presencia de mujeres que dicen haber conocido gente, actualizado conocimientos o otros motivos personales, y leve mayoría de hombres que dicen no haber logrado ningún objetivo todavía gracias a la actividad educativa.

Figura I-19. Objetivos a los que ha contribuido la actividad



Fuente: Elaboración propia

Clustering

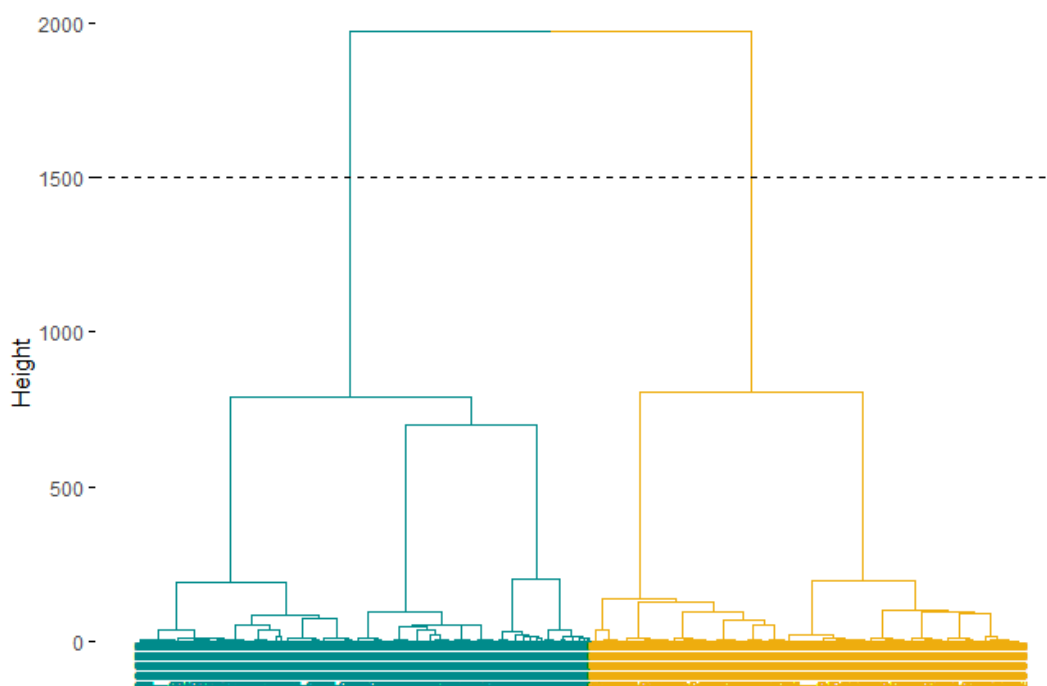
A continuación, para cerrar este primer bloque, se construirán diversos *clústers* en torno a diferentes actitudes o características estudiantiles, siempre analizando si el género es determinante en estas agrupaciones.

Se calcularán las agrupaciones con las **variables generales** del nivel de estudios, el contenido de estos y si se ha abandonado alguna otra educación.

En primer lugar, se añade el sexo en el clúster, a modo confirmatorio de las características vistas para los grupos de mujeres y hombres.

Siguiendo la metodología explicada en el apartado metodológico, se obtienen dos clústers donde el 89.9% del clúster 1 son hombres y el 99.89% del clúster 2, mujeres, con lo que se podría determinar que cada clúster representa el perfil de los individuos de cada género. Con esto se presenta el siguiente dendrograma donde se diferencian dos claros clústers: hombres en amarillo y mujeres en azul.

Figura I-20. Dendrograma clústering jerárquico variables generales y género



Fuente: Elaboración propia

Con todo esto, se obtienen los resultados de R en el Anexo 4, mediante los cuales se realiza un perfilado de cada clúster, comentando aquellas categorías que se encuentran sobrerrepresentadas en cada uno y, por lo tanto, que se asocian en mayor medida a hombres o a mujeres.

El 99.7% de aquellos que dicen haber abandonado una actividad formativa se encuentran en el **primer clúster**, con lo que se podría relacionar mayoritariamente con los hombres. También en este primero, se encuentran el 89.71%, el 75.14% y el 74.63% de aquellos que han estudiado en ámbitos de mecánica, electrónica, industria y construcción, en las TIC o en agricultura y ganadería, respectivamente. En cuanto a los niveles de estudio, encontraríamos mayoría en el clúster 1 de aquellos con máximo nivel formativo de formación profesional de grado superior (59.74%), bachillerato (53.58%), formación profesional (53.79%).

Por otro lado, analizando el **segundo clúster**, asociado a las características de las mujeres, se obtiene lo siguiente. En cuanto al contenido de los estudios, se encuentra una mayor variedad de categorías sobrerrepresentadas que en el primer clúster; el 71.94% de los individuos dedicados a la salud y los servicios sociales se encuentran en este primer clúster. También el 60% y el 68% de los que se estudiaron negocios, administración y derecho o educación, respectivamente, se encontrarían en este clúster. En referencia a nivel de estudios, habría

mayoría de grados universitarios, diplomaturas y licenciaturas y másteres y doctorados, con un 57.43%, 52.35% y 52.8% de los individuos con estos niveles presentes en el clúster 2. También se encuentran el 56.69% de aquellos que dicen no haber abandonado una actividad educativa.

Con esto, se obtiene la siguiente tabla, donde se ven representadas las características más claras asociadas a cada clúster y también a cada género, siendo estos mayoría en cada uno de los clústers.

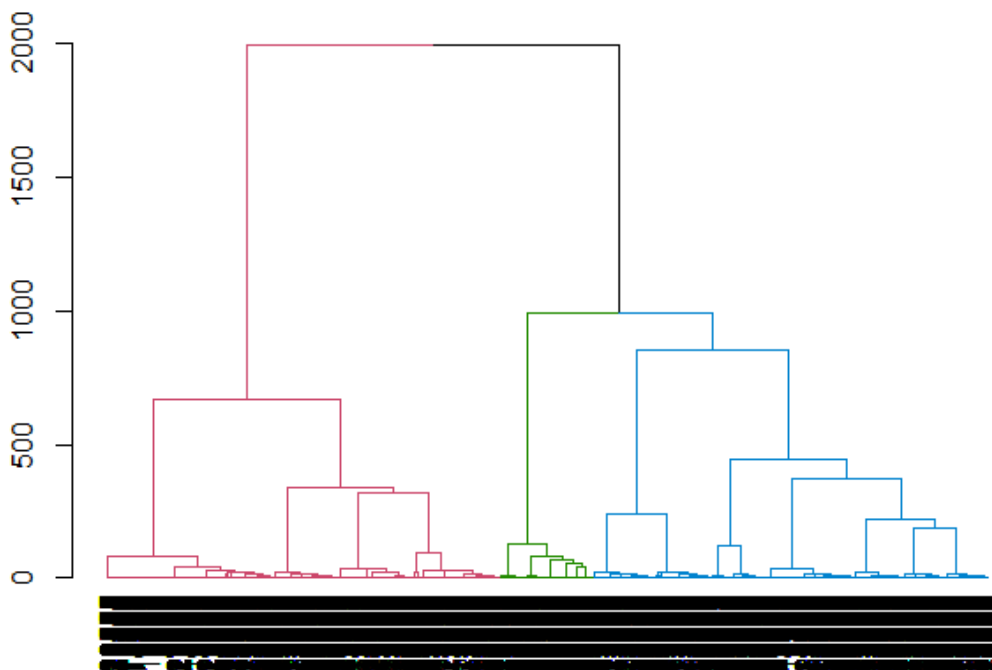
Tabla I-3. Perfilado clústers variables generales y género

ÁREAS	HOMBRES	MUJERES
Temáticas	✓ Profesiones manuales ⁴ ✓ TIC	✓ Salud y servicios sociales ✓ Educación
Nivel de estudios	✓ Nivel intermedio	✓ Nivel superior
Abandono educación	✓ Sí	✓ No

Fuente: Elaboración propia

Por otro lado, con estas mismas variables, se calcularán los clústers con las mismas variables, pero sin utilizar la variable de género, de esta manera, se quiere ver como esta interactúa en los diferentes clústers sin forzarlo. Con esto quedaría más claro en que variables o categorías se encuentran las mayores diferencias entre hombres y mujeres.

Figura I-21. Dendrograma clústering jerárquico variables generales sin género



Fuente: Elaboración propia

⁴ Hace referencia a mecánica, electrónica, industria y construcción y a agricultura, ganadería, pesca, silvicultura y veterinaria

Siguiendo los resultados obtenidos en el Anexo 5, se realiza el perfilado de los clústers.

Tabla I-4. Perfilado clústers variables generales sin género

ÁREAS	CLÚSTER 1	CLÚSTER 2	CLÚSTER 3
Género	Hombres (48%)	Mujeres (48.5%)	Hombres (11.83%)
Temáticas	- Formación general - Profesiones manuales - Ciencias naturales	- Educación - Salud - Ciencias sociales - Negocios	- TIC - Servicios
Nivel estudio	- Bachillerato, FP - Grado, licenciatura	- Grado, diplomatura - Máster, doctorado	- Formación profesional
Abandono	No aplicable	No	Sí

Fuente: Elaboración propia

En el **clúster 1**, con el 48% de los hombres, se pueden ver reflejadas algunas de las características ya asociadas con este género, como serían las profesiones manuales o los niveles de estudios intermedios. En el **clúster 2**, con mayor presencia de mujeres (el 48.5% de ellas), también se observan estas características "típicas" como son los ámbitos de la educación o la salud, los niveles de estudio de grado o máster o el no abandono de actividades educativas. En el **clúster 3**, de la misma manera, con sobrerrepresentación masculina, también saldrían algunas de estas características asociadas con ellos, como el área de las TIC, la formación profesional como nivel de estudios o el abandono de alguna actividad educativa.

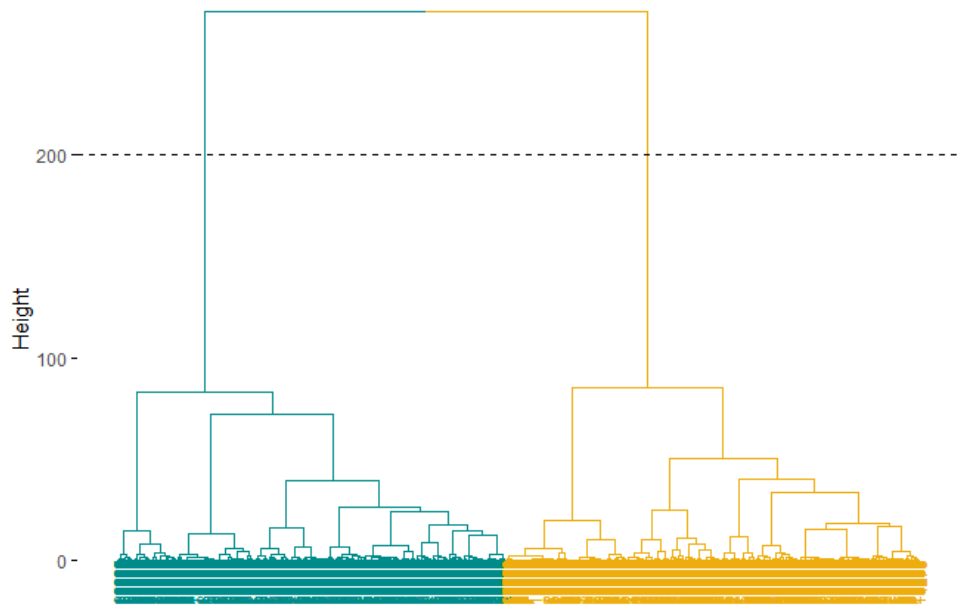
Con estos clústers, se siguen percibiendo algunas de las características claras asociadas a cada género, aunque confundidas entre otras no tan claras. Es por esto que sería más determinante y vinculante el cálculo de los clústers añadiendo la variable género, con la intención de centrar la atención en aquellas variables más presentes en cada género.

A continuación, se construyen clústers con los **actuales estudiantes**, con la intención de caracterizar los perfiles de estos. Basando el cálculo en los resultados anteriores, las agrupaciones de calcularan añadiendo la variable del género, con la intención de obtener de manera clara aquellas características relacionadas, en mayor medida, a mujeres y hombres.

En cuanto al contenido, se utilizarán las variables del número de actividades educativas realizadas, el nivel de estudios cursado, la temática de los estudios, si estos han sido realizados

en horas de trabajo y quien paga la actividad en cuestión. El dendrograma obtenido es el siguiente.

Figura I-22. Dendrograma clústering jerárquico estudiantes actuales



Fuente: Elaboración propia

De nuevo se representan los clústers 1 y 2 como hombres y mujeres, puesto que en primero se encuentran el 92% de las mujeres y en el segundo el 95% de los hombres.

Con los resultados del Anexo 6, se realiza el siguiente perfilado de cada uno de los clústers.

Tabla I-5. Perfilado clústers estudiantes actuales

ÁREAS	HOMBRES	MUJERES
Temáticas	<ul style="list-style-type: none"> ✓ Profesiones manuales ✓ Formación general ✓ TIC 	<ul style="list-style-type: none"> ✓ Salud y servicios sociales ✓ Educación ✓ Artes, humanidades
Nivel de estudios	<ul style="list-style-type: none"> ✓ Educación secundaria ✓ Educación primaria 	<ul style="list-style-type: none"> ✓ Grado, licenciatura, diplom. ✓ Máster, doctorado
Pago estudios	<ul style="list-style-type: none"> ✓ Actividad gratuita ✓ Pago total por otra persona/institución 	<ul style="list-style-type: none"> ✓ Pagado totalmente por usted
En horas de trabajo	<ul style="list-style-type: none"> ✓ Mayoritariamente o al completo en horas de trabajo remuneradas 	<ul style="list-style-type: none"> ✓ Fuera de horas de trabajo remuneradas

Fuente: Elaboración propia

En estos otros clústers sobre los actuales estudiantes, se pueden seguir viendo características históricas ligadas a los géneros, como serían los contenidos o los niveles de estudio, ellos

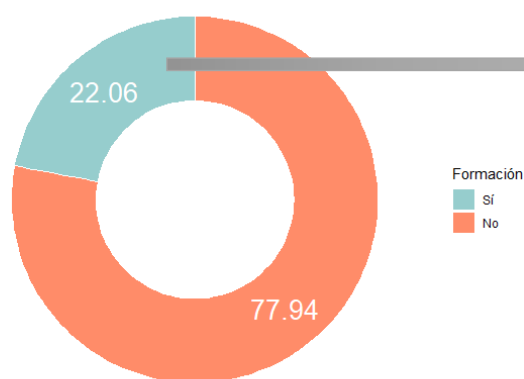
estudian contenidos de nivel medio relacionados con profesionales manuales o las TIC; ellas obtienen mayores niveles de estudios y en sectores como la salud, la educación o las humanidades y las artes. Analizando las nuevas variables que se recogen para los estudiantes actuales, se perciben diferencias clave entre hombres y mujeres, ya observadas a lo largo del trabajo. Los hombres son mayoría entre aquellos que participan en actividades gratuitas o bien estas son pagadas por otra persona o institución, mientras que entre las mujeres destaca el pago propio de la actividad educativa. También en referencia a si la actividad se ha llevado a cabo durante horas de trabajo remuneradas, entre ellas destaca la realización de actividades fuera de horas de trabajo, mientras que entre ellos destaca la realización mayoritaria o total de la actividad en horas de trabajo remuneradas.

1.2. Educación no formal

Una vez analizados los aspectos educativos formales de los individuos, se pasa al ámbito de educación no formal⁵.

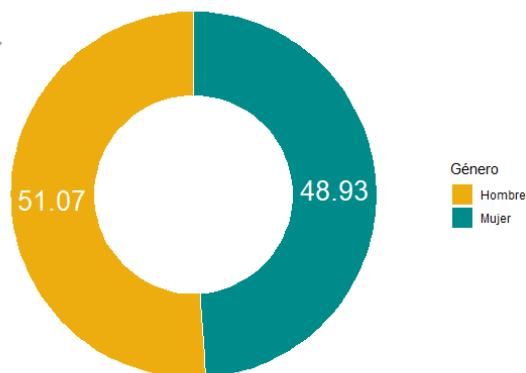
En este caso, las variables recogen si se ha realizado o se está realizando formación en el puesto de trabajo, de que tipo, la dedicación a estas y diversos detalles más; también se observan variables como posibles dificultades para el acceso a formación, percepción sobre la educación propia recibida y diversas variables sobre aprendizaje autónomo o informal.

Figura I-24. Formación en los últimos 12 meses



Fuente: Elaboración propia

Figura I-23. Presencia de géneros en formación actual (base: 5079)



Fuente: Elaboración propia

⁵ Se considera educación no formal aquellas actividades destinadas a aumentar conocimientos o habilidades con ayuda de un profesor o instructor, por las que no se conseguirá un título reconocido por el Ministerio o Consejería de Educación

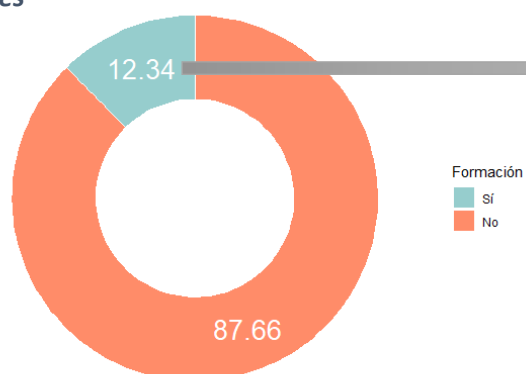
En los últimos 12 meses un 22% de la población ha participado o está participando en alguna actividad de formación planificada para su puesto de trabajo organizada por su empleador.

De estos, un 52.421% son hombres y el 47.49% restante, mujeres. Produciéndose entre ambos sexos una diferencia de casi 5 puntos. El test realizado, indica que existen diferencias entre hombres y mujeres en cuanto a la participación, o no, en actividades formativas en el entorno laboral.

Con esto se concluye, que son los hombres quienes reciben mayor formación en su puesto de trabajo.

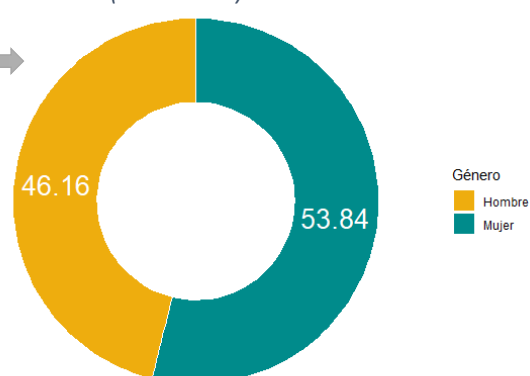
Por otro lado, si se observa la participación en cursos, talleres o seminarios⁶, fuera del ámbito laboral, se obtienen los siguientes resultados.

Figura I-26. Curso o seminario en los últimos 12 meses



Fuente: Elaboración propia

Figura I-25. Presencia de géneros en cursos y seminarios (base: 2840)



Fuente: Elaboración propia

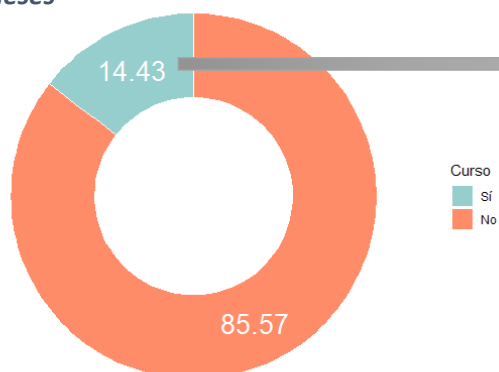
Del 12.34% de personas que participan o han participado en actividades de este tipo, un 53.84% son mujeres.

En este caso, también se producen diferencias entre sexos, siendo ellas las que más participan en cursos o seminarios fuera del entorno laboral.

⁶ Se contemplan talleres de empleo, obtención del permiso de conducir, cursos de informática, cocina, etc.

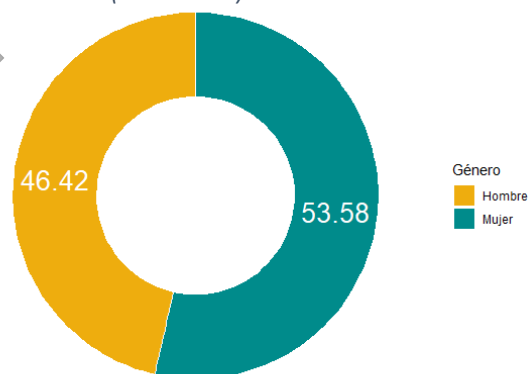
De la misma manera, se contemplan cursos teóricos, charlas o conferencias.

Figura I-28. Charla o conferencia en los últimos 12 meses



Fuente: Elaboración propia

Figura I-27. Presencia de géneros en charlas o conferencias (base: 3322)

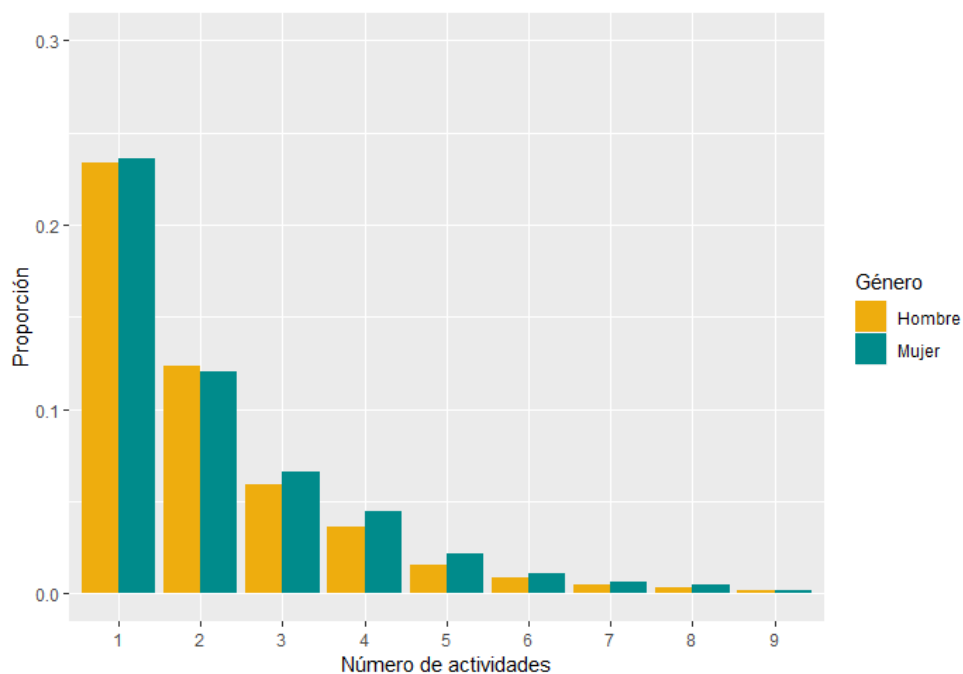


Fuente: Elaboración propia

De nuevo, de los que dicen haber asistido a charlas o conferencias, un 53.58% pertenecen al género femenino. En este caso, las diferencias entre sexos no resultan significativas, con lo que no se podría determinar que mujeres y hombres tengan una actitud diferente en torno a la asistencia a charlas o conferencias.

Con todo esto, se recogen el número de actividades de educación no formal a las que se ha asistido en los últimos 12 meses, contando formaciones en el entorno laboral, cursos, seminarios, charlas o conferencias.

Figura I-29. Número de actividades de educación y formación no formal (base:



* se han omitido las respuestas de más de 9 actividades, que contenían 30 casos en total NFENUM

** se contemplan también clases particulares que no han sido estudiadas al detalle

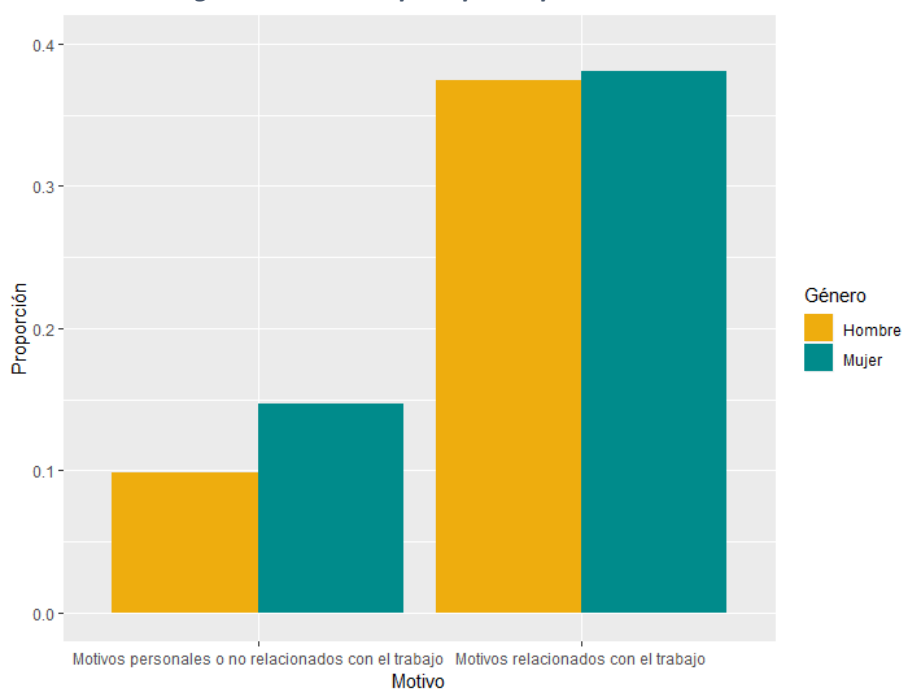
Fuente: Elaboración propia

A priori, no se perciben grandes diferencias entre géneros en cuanto al número de actividades realizadas, aunque con la realización del test, se determina que existen diferencias.

En relación a la participación en actividades, hombre y mujeres no tienen un mismo comportamiento, siendo ellas las que participan en más actividades.

Siguiendo con este tipo de actividades, se estudian variables de las características concretas⁷ sobre las actividades no formales realizadas.

Figura I-30. Razón para participar en la actividad



Fuente: Elaboración propia

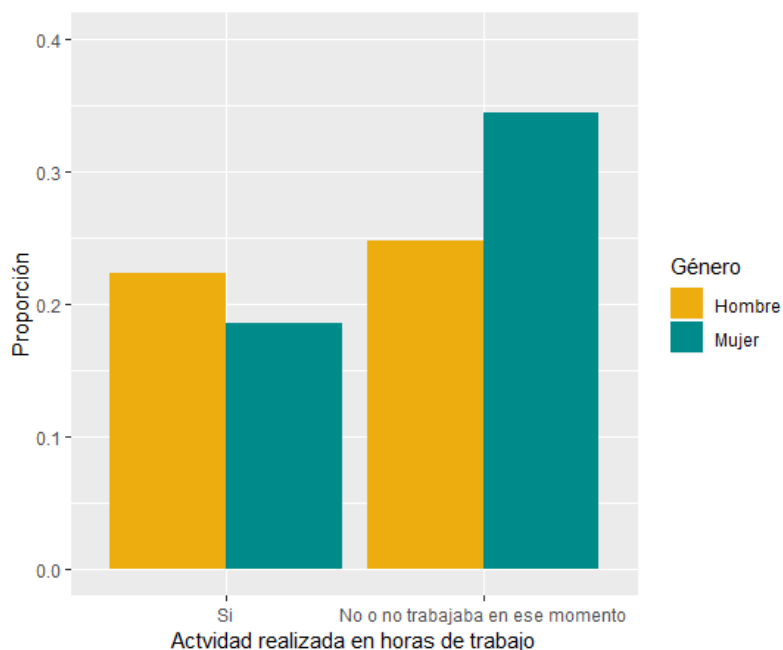
Sobre la razón principal que lleva a los individuos a participar en una determinada actividad, se observa cómo más mujeres destacan motivos personales no relacionados con el trabajo.

Estas diferencias resultan significativas, lo que determinaría que existen diferencias en torno a las razones que llevan a hombres y mujeres a participar en actividades educativas no formales, siendo ellas las que tienen motivaciones más de tipo personal y ellos más relacionadas con el trabajo.

⁷ Para analizarlas, se han recogido las respuestas válidas sobre las 7 posibles actividades no formales

Cuando se observa si la actividad en cuestión ha sido realizada en horas de trabajo, ya se percibe algo parecido a lo observado con otras variables, siendo los hombres quienes más realizan actividades educativas en horas de trabajo.

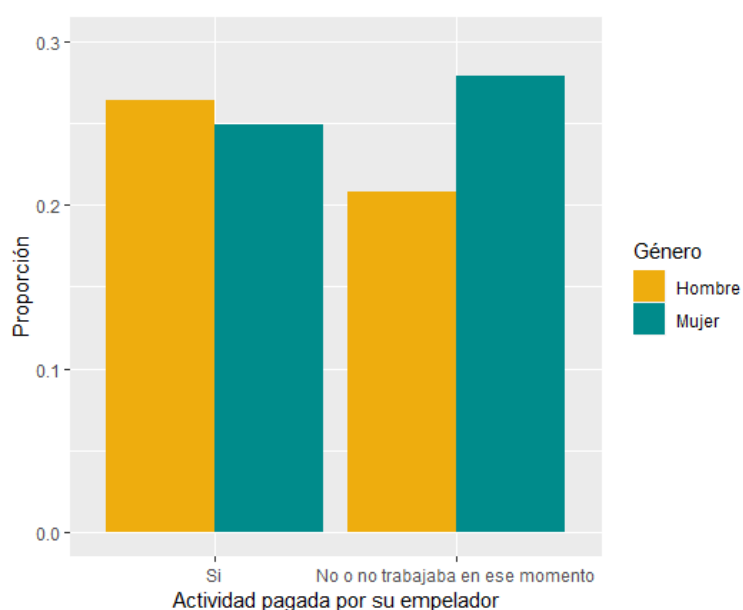
Figura I-31. Actividad realizada en horas de trabajo



Fuente: Elaboración propia

También en relación con esta variable se observa una mayor presencia de hombres entre aquellos a los que el empleador paga la actividad en cuestión.

Figura I-32. Actividad pagada por el empleador



Fuente: Elaboración propia

Con estas dos, se extrae la idea de que son los hombres quienes más actividades formativas realizan en sus puestos de trabajo y, en consecuencia, a quienes más pagan actividades las empresas.

Por otro lado, entrando al detalle de las razones que han motivado la realización de la actividad, se vuelve a percibir visualmente una mayor presencia de mujeres en categorías como “conocer gente / diversión” o “adquisición de conocimientos y habilidades útiles para la vida/ de interés”, mientras que los hombres se encontrarían más presentes en opciones como “crear un negocio propio”.

Figura I-33. Razón o razones para participar en la actividad



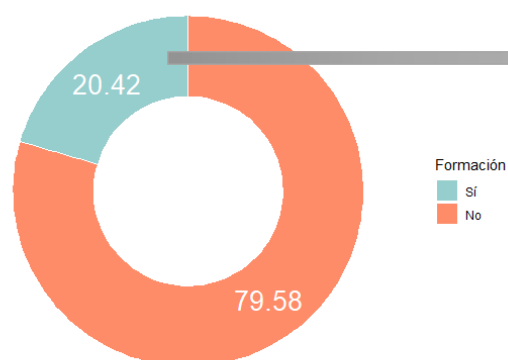
Fuente: Elaboración propia

Estos matices, siguen reforzando la idea de que las mujeres tienen motivaciones más personales y no tan laborales como los hombres, quienes demuestran una actitud más enfocada al negocio y hacia el crecimiento profesional.

También se estudian las dificultades que hayan podido encontrar los individuos a la hora de participar en actividades educativas y formativas.

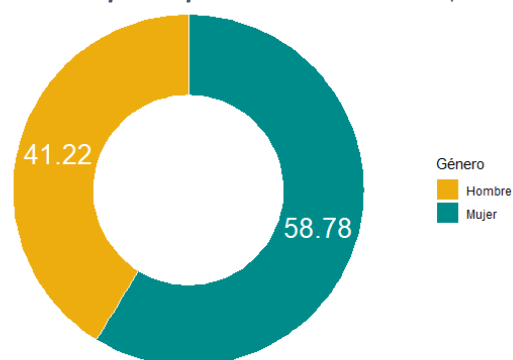
Por una parte, se observan aquellas personas que están estudiando o han estudiado en los últimos 12 meses, ya sea educación formal o no formal.

Figura I-34. Intención de participar en alguna otra actividad educativa (base: 10096)



Fuente: Elaboración propia

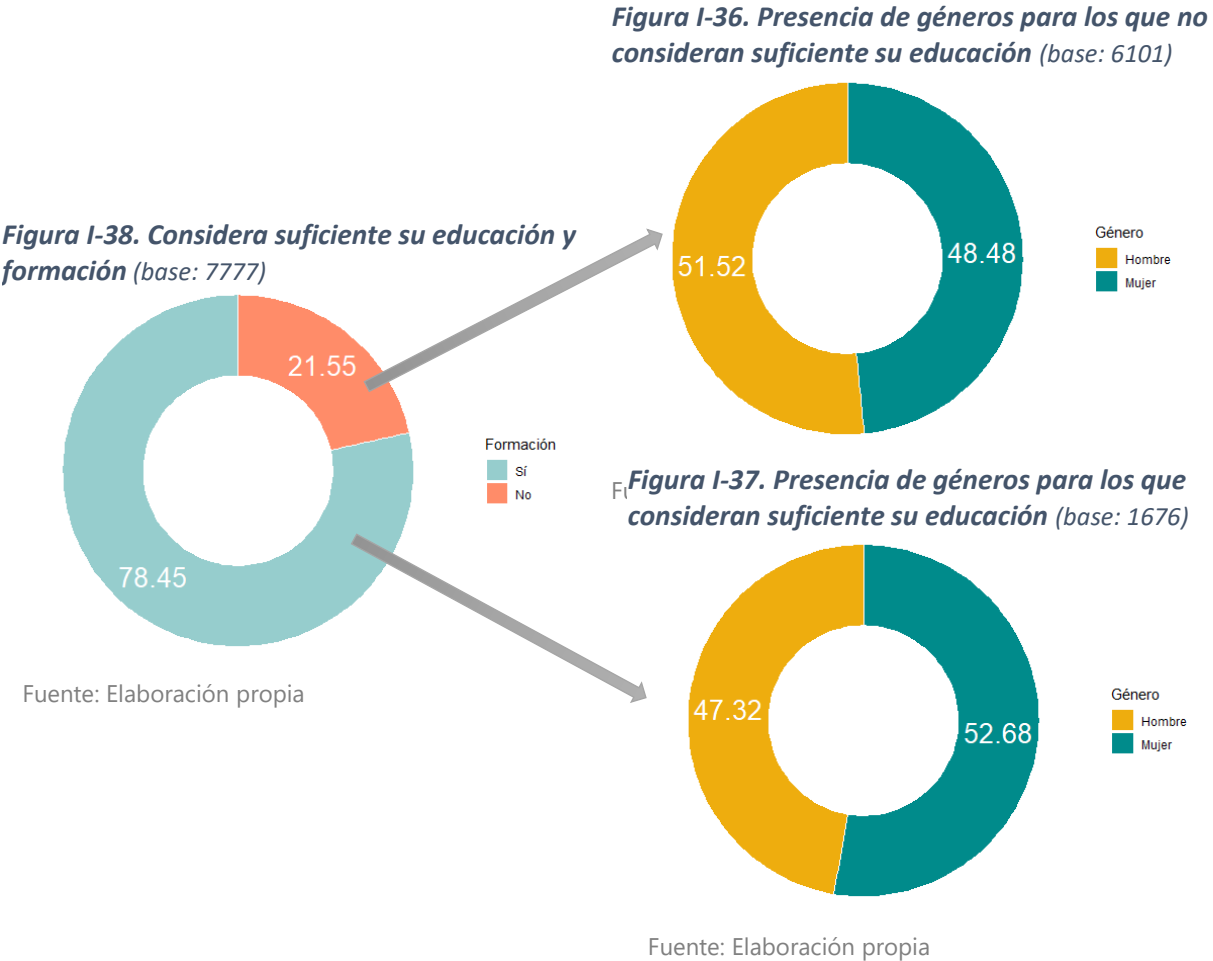
Figura I-35. Presencia de géneros en la intención de participar en otra actividad (base: 10096)



Fuente: Elaboración propia

Un 20.42% de estos quiso participar en otras actividades extra en el mismo período en cuestión, de los cuales un 58.78% son mujeres y el 41.22% restante, hombres. Las diferencias entre estos resultan significativas, con lo que se podría determinar que son ellas las que tienen una mayor intención de participar en actividades educativas.

Posteriormente, si se centra la atención en el 79.58% de personas que no han tenido intención de participar en alguna otra actividad extra en los últimos doce meses, a estos se les cuestiona sobre si consideran suficiente la educación o formación que ha estado o está cursando.

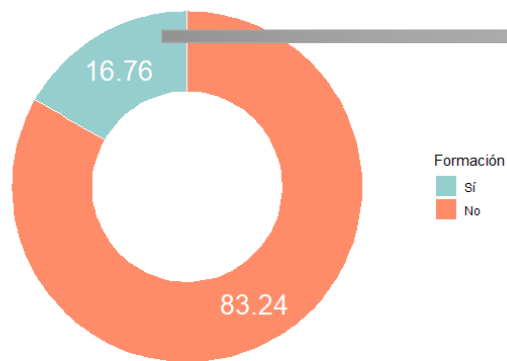


Dentro de aquellos que no han realizado actividades extra en los últimos 12 meses y tampoco han tenido intención, se encuentra un 78.45% que ya considera suficiente la educación y formación cursada, contra un 21.55% que no la considera suficiente.

Entre los que ya consideran suficiente la educación y formación en la que se ha participado, se encuentra, significativamente, mayoría de mujeres; mientras que entre los individuos que no consideran suficiente esta educación, se encontraría una presencia significativamente mayor de hombres.

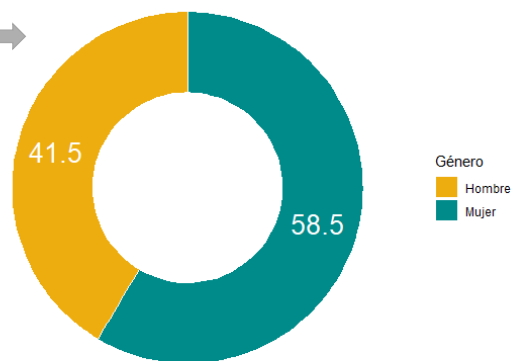
Por otro lado, se les pregunta sobre el deseo de participar en alguna actividad de educación y formación, a aquellos que no han participado en ninguna actividad de este tipo en los últimos 12 meses.

Figura I-40. Intención de participar en alguna actividad educativa (base: 12206)



Fuente: Elaboración propia

Figura I-39. Presencia de géneros en la intención de participar en alguna actividad educativa (base: 2046)

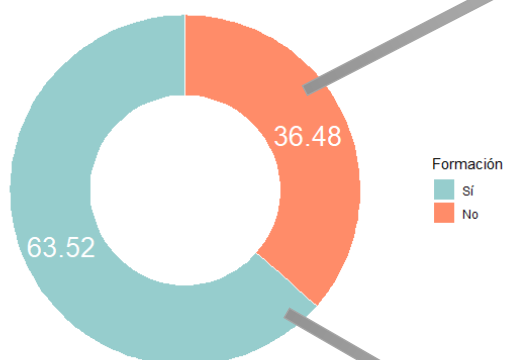


Fuente: Elaboración propia

Se obtienen resultados muy parecidos a los observados anteriormente, donde son las mujeres las que tienen una mayor presencia, significativa, entre aquellos que si han querido participar en actividades formativas o educativas.

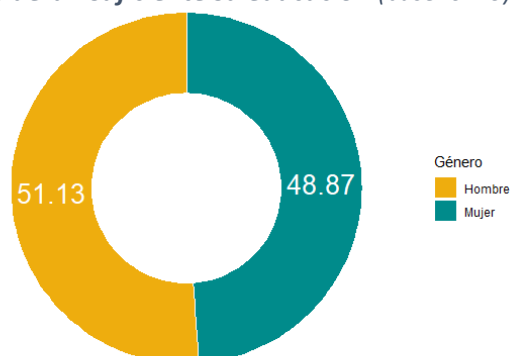
Posteriormente, para aquellos que dicen no haber tenido intención de participar en actividades educativas, hay un 63.52% que ya considera suficiente su educación y formación, versus un 36.48% que no la consideran suficiente.

Figura I-43. Considera suficiente su educación y formación (base: 9800)



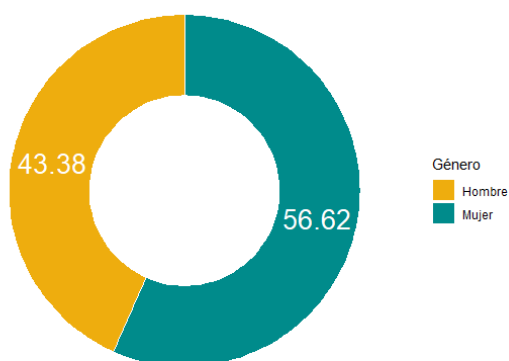
Fuente: Elaboración propia

Figura I-41. Presencia de géneros para los que no consideran suficiente su educación (base: 6225)



Fuente: Elaboración propia

Figura I-42. Presencia de géneros para los que consideran suficiente su educación (base: 3575)

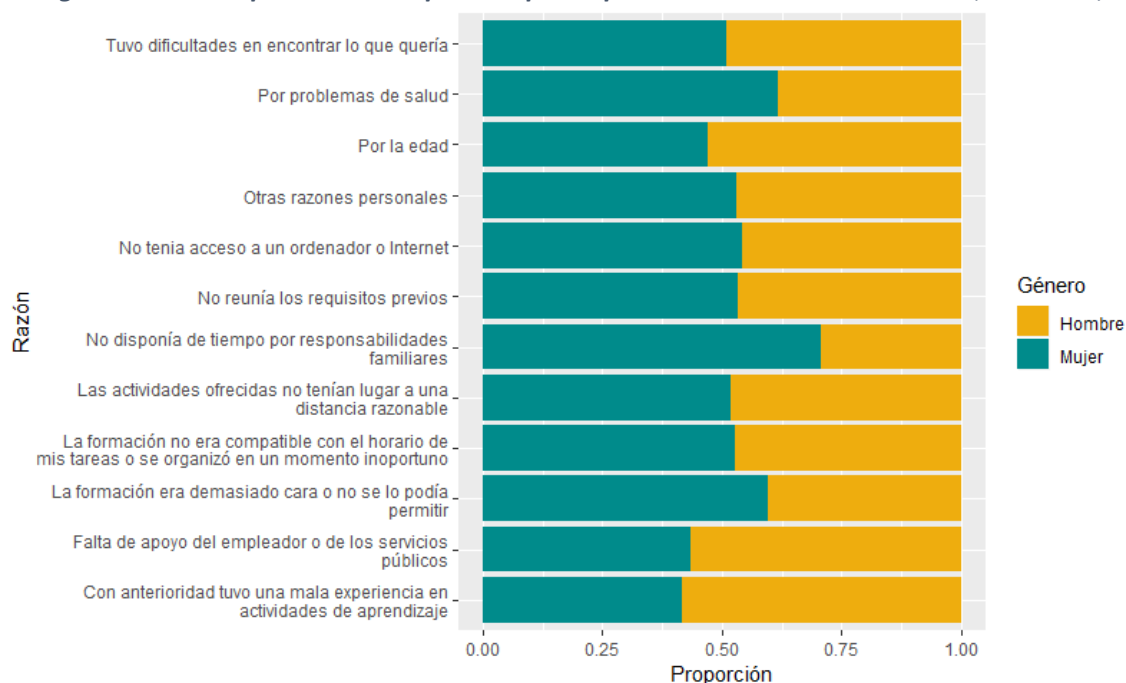


Fuente: Elaboración propia

De manera reiterada, se obtienen resultados similares a los ya comentados anteriormente. Representan las mujeres una mayoría significativa entre los que ya consideran suficiente su educación y formación, mientras que los hombres la representan entre los que no consideran suficiente sus conocimientos.

En último lugar, en el entorno de las actividades educativas no formales, se estudian los principales motivos que los individuos destacan como dificultades para la participación en actividades de educación y formación.

Figura I-44. Principales motivos para no participar en actividades educativas (base: 8553)



Fuente: Elaboración propia

De estos motivos, las mujeres serían mayoría en tres de ellas, siendo las demás más dichas por los hombres. Ellas destacan problemas de salud, falta de tiempo por responsabilidades familiares o dificultades económicas para acceder a la formación.

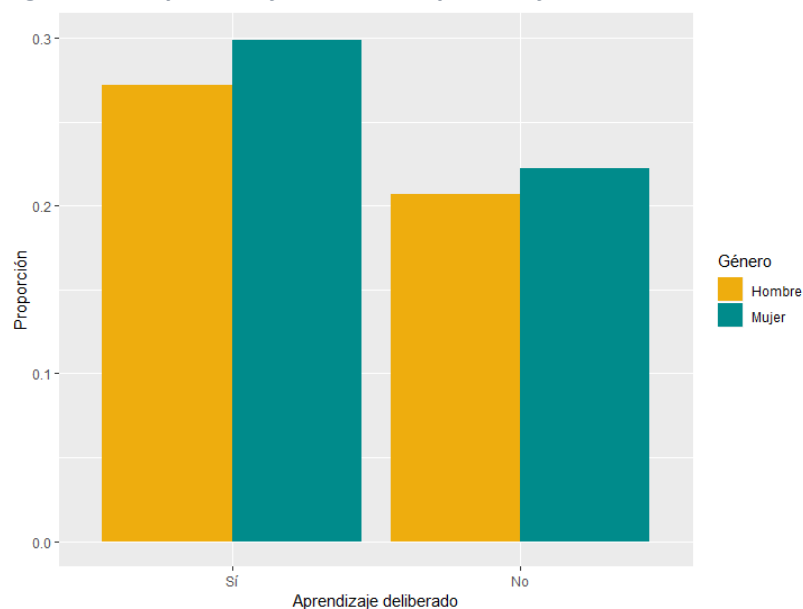
Estos tres motivos denotan una actitud históricamente ligada a las mujeres, de anteposición de la vida familiar al crecimiento personal, y esto se sigue notando mediante esta variable, donde ellas no destacan dificultades ligadas al deseo o a impedimentos externos, sino que estas son más ligadas al entorno familiar.

1.3. *Aprendizaje informal*

Finalmente, para cerrar el análisis de la educación, se estudiarán variables sobre aprendizaje informal.

A continuación, se recoge la variable sobre si se han llevado actividades para mejorar

Figura I-45. Aprendizaje deliberado para mejorar conocimientos

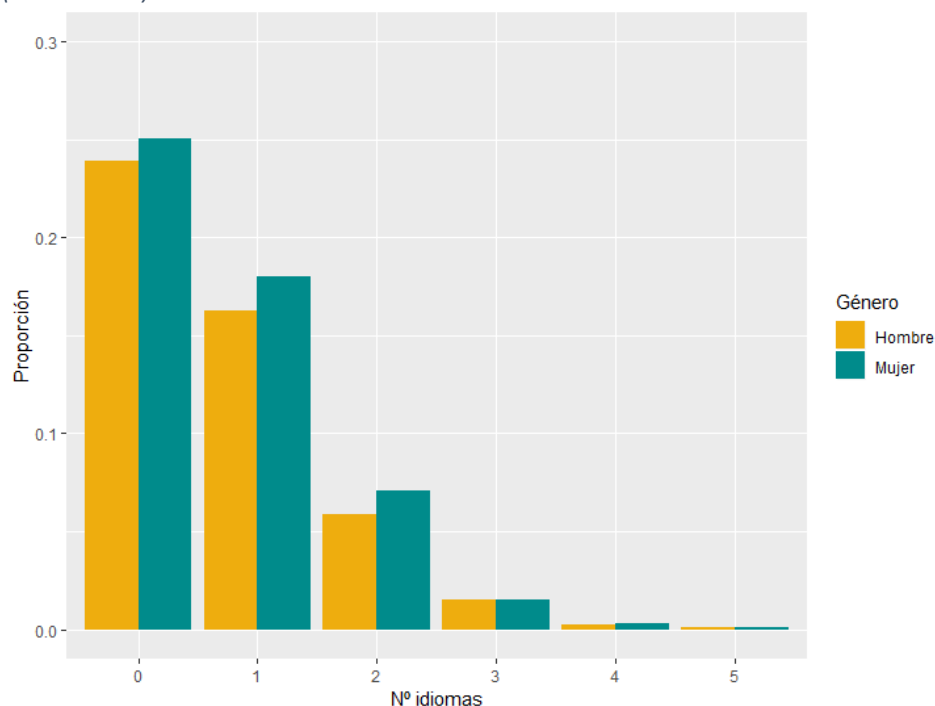


conocimientos o habilidades, donde se contemplan actividades por medio de familiares o amigos, libros, ordenadores, etc. En primer lugar, se percibe una mayor proporción de individuos en el sí; cuando estos se diferencian según su género, no se perciben diferencias significativas. Se concluye que no hay diferencias de género en torno a las actividades de aprendizaje informal.

Fuente: Elaboración propia

Finalmente, se contempla el número de idiomas que hablan los individuos distintos de su o sus lenguas maternas.

Figura I-46. Conocimiento de idiomas distintos de su/sus lenguas maternas
(base: 23010)



* se han excluido 9 casos de individuos que hablaban 6, 7, 8 o 15 idiomas

Fuente: Elaboración propia

Se puede percibir visualmente una mayor presencia de mujeres en todos los grupos, por lo tanto, para saber si estas diferencias son significativas o no se realiza un contraste de hipótesis. De este se obtiene que existen diferencias significativas⁸, siendo las mujeres quienes saben un mayor número de idiomas.

Con esto se cerraría el capítulo de educación, de donde se han podido extraer las siguientes afirmaciones.

Las mujeres se sitúan en niveles educativos extremos, a menor edad estas se concentran en altos niveles educativos, mientras que cuanto más mayores, menor nivel educativo tienen; por consiguiente, las categorías educativas donde habría mayor concentración de hombres sería en los niveles intermedios. Sigue habiendo claras temáticas de hombres y de mujeres, produciéndose el mismo fenómeno tanto históricamente (personas de mayor edad) como entre los estudiantes actuales. Los hombres son mayoría entre aquellos que abandonan una enseñanza, así como entre aquellos que realizan la actividad educativa en horas de trabajo y,

⁸ El test resulta significativo tanto excluyendo los casos extremos como contemplándolos.

en consecuencia, mayoría entre las personas a quienes las empresas pagan la educación; por otra parte, las mujeres tienen mayor presencia entre las personas que se pagan la educación de manera independiente.

Por otro lado, no solo se han encontrado diferencias, siendo hombres y mujeres iguales ante el número de actividades formativas formales que realizan y en torno a la consideración de frecuencia de uso de los conocimientos adquiridos mediante la actividad educativa en curso.

En cuanto a la educación no formal, es decir, aquella que se produce a lo largo de la carrera profesional de las personas y no solo en la etapa estudiantil, se han observado diversos fenómenos. Los hombres realizan más actividades formativas en sus puestos de trabajo, en cambio, fuera del entorno laboral, son ellas las que más participan en cursos y seminarios. En general, recogiendo todas las actividades educativas y formativas no formales, ellas participan en un mayor número de actividades teniendo motivaciones de tipo personal y realizándolas fuera del trabajo; por otra parte, ellos presentan motivaciones relacionadas con el entorno laboral, llevando las actividades a cabo en horas de trabajo y, por consiguiente, siendo el empleador el que paga dicha actividad.

En torno a las dificultades para participar en actividades educativas también se han detectado diferencias. Ellas presentan una mayor intención de participar en actividades educativas, hayan estudiado o no en los últimos 12 meses. Por otro lado, dentro de aquellos que no han tenido intención, cuando se les pregunta si consideran suficiente su educación, se encuentra mayoría de hombres entre quienes no la consideran suficiente y ellas son mayoría entre los que si consideran suficiente su formación.

Finalmente, también se observa un mayor conocimiento de idiomas por parte de las mujeres.

Capítulo II. ¿Brecha en inserción laboral?

En este segundo capítulo, se explorará el momento posterior a la educación, ya estudiada en el apartado anterior, como es la inserción laboral.

Vistos puntos clave de diferenciación entre hombres y mujeres en el punto de partida de la carrera profesional, con estos datos se buscará apoyar y reforzar el análisis, viendo como estas brechas iniciales pueden transcurrir durante la vida profesional de las personas, en este caso en el punto concreto de la inserción laboral. Estos datos, pero, se tratarán para los estudiantes universitarios, partiendo de la idea de que este sector estudiantil es o debería ser donde se producen menores diferencias de género.

Quizá el punto más importante de relación entre estos los dos primeros capítulos sería el sector al que se dedican los individuos. Se cree que es un factor muy determinante en las diferencias que puedan ocasionarse posteriormente a la educación. Es por esto, con la intención de obtener resultados factibles para estudiar este fenómeno, que se ha realizado una elección de sectores que se estudiarán en las siguientes páginas. El proceso llevado a cabo se encuentra en los anexos de Preprocessing.

Como ya se ha ido comentando, en este capítulo se tratará de analizar el paso del mundo educativo al mundo laboral, y como se comportan las posibles diferencias entre géneros. Esto complementado con la aplicación de una técnica no empleada hasta el momento, que son las reglas de asociación. Con la intención de abordar unos datos con la misma finalidad que en el primer capítulo, pero con diferente metodología, se irán analizando grupos de variables por áreas temáticas, tratando de encontrar patrones de comportamiento mediante las reglas de asociación y observando también si en estas interactúan los géneros de diferente manera, o no. Todo esto también se complementará con análisis cruzados y contrastes de hipótesis cuando sea necesario.

A continuación, se ha llevado a cabo el cálculo de las principales reglas en los datos mediante el método *a priori*. Como ya se comentaba en el apartado metodológico, se considerarán aquellas reglas con las categorías de hombre o mujer implicadas y con un *lift* mayor a 1, es decir, aquellas donde no se haya podido afirmar la independencia entre categorías y, por lo tanto, haya dependencia entre el género y la categoría en cuestión. Para el mayor entendimiento analítico de las reglas, solo se presentarán aquellas con la categoría género y otra implicada, sin considerar reglas con 3 o más categorías.

A continuación, se presenta una tabla con las variables implicadas en las reglas obtenidas, para mayor entendimiento de estas.

Tabla II-1. Definición de las variables que aparecen en las reglas de asociación

VARIABLE	DEFINICIÓN
JORNADA	Tipo de jornada en el trabajo actual
TRAB_F1	Trabajo remunerado durante la carrera (sin ser prácticas)
P_CONOC	Uso en su primer empleo de los conocimientos y habilidades adquiridos en la carrera
P_JOR	Tipo de jornada de trabajo en su primer empleo
BQ_TP_T	Tiempo que ha estado trabajando desde que terminó los estudios
BSQCOM2	Búsqueda de primer empleo: a través de servicios de empleo públicos o de la universidad
BSQCOM7	Búsqueda de primer empleo: preparó unas oposiciones
BSQCOM3	Búsqueda de primer empleo: a través de empresas de trabajo temporal
PSIT_PR	Situación profesional de su primer empleo
BSQCOM4	Búsqueda de primer empleo: contacto con el empleador por iniciativa propia o contactos personales
BSQCOM1	Búsqueda de primer empleo: anuncios en el periódico, internet
BSQRE	Rechazo de algún trabajo por considerarlo poco adecuado desde que se ha titulado
PRAC	Realización de prácticas en empresas
BSQCOM9	Búsqueda de primer empleo: de otra forma
BSQCOM5	Búsqueda de primer empleo: empleador se puso en contacto con la persona
BSQCOM6	Búsqueda de primer empleo: continuación de las prácticas realizadas
BSQCOM8	Búsqueda de primer empleo: creación de un negocio propio

Una vez definidas las variables implicadas, se presentan las reglas obtenidas con las condiciones mencionadas anteriormente.

Figura II-1. Reglas de asociación obtenidas mediante el método apriori.

	lhs	rhs	support	confidence	lift	count
[1]	{SEXO=Hombre}	=> {JORNADA=A tiempo completo}	0.2737141	0.6549667	1.127652	5598
[2]	{SEXO=Hombre}	=> {TRAB_F1=Si}	0.2617837	0.6264186	1.053145	5354
[3]	{SEXO=Hombre}	=> {P_CONOC=Si}	0.2539605	0.6076986	1.006532	5194
[4]	{SEXO=Hombre}	=> {P_JOR=A tiempo completo}	0.2877469	0.6885457	1.098536	5885
[5]	{SEXO=Hombre}	=> {BQ_TP_T=2 o más años}	0.2805105	0.6712297	1.036936	5737
[6]	{SEXO=Hombre}	=> {BSQCOM2=No}	0.2915607	0.6976717	1.023365	5963
[7]	{SEXO=Hombre}	=> {BSQCOM7=No}	0.3258850	0.7798058	1.021821	6665
[8]	{SEXO=Hombre}	=> {BSQCOM3=No}	0.3286720	0.7864748	1.000185	6722
[9]	{PSIT_PR=Asalariado con contrato temporal}	=> {SEXO=Mujer}	0.2904361	0.6572250	1.129069	5940
[10]	{BSQCOM4=No}	=> {SEXO=Mujer}	0.3232936	0.5852363	1.005397	6612
[11]	{SEXO=Mujer}	=> {BSQCOM1=No}	0.3503814	0.6019320	1.021043	7166
[12]	{SEXO=Mujer}	=> {BSQRE=No}	0.3785938	0.6503990	1.070322	7743
[13]	{SEXO=Mujer}	=> {PRAC=Si}	0.4340407	0.7456531	1.090461	8877
[14]	{SEXO=Mujer}	=> {BSQCOM9=No}	0.4367299	0.7502730	1.000364	8932
[15]	{SEXO=Mujer}	=> {BSQCOM5=No}	0.4478780	0.7694246	1.014654	9160
[16]	{SEXO=Mujer}	=> {BSQCOM6=No}	0.4765304	0.8186476	1.011477	9746
[17]	{SEXO=Mujer}	=> {BSQCOM8=No}	0.4996088	0.8582948	1.021344	10218

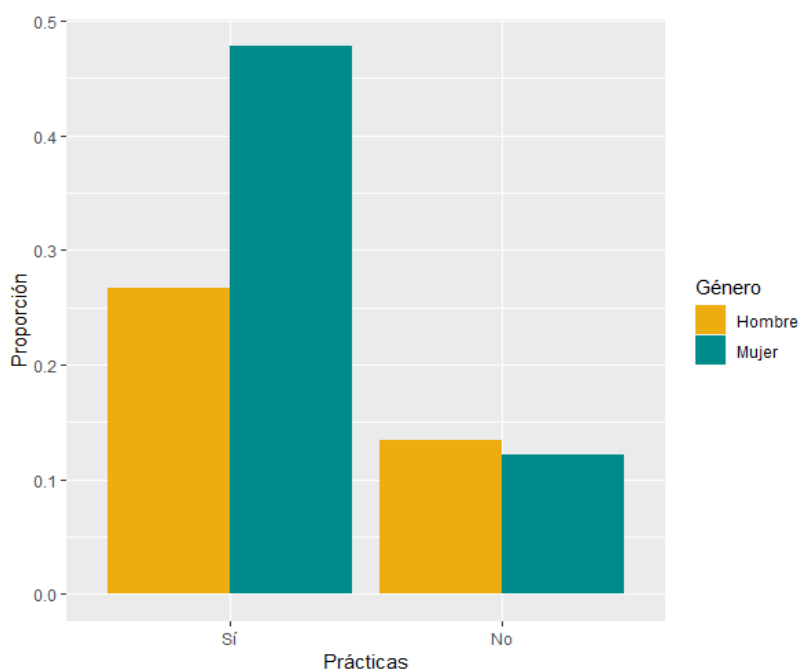
A continuación, mediante la estructuración en bloques temáticos, se irán mencionando los resultados o ideas extraídas de las reglas, complementando cuando sea necesario con gráficos o contrastes de hipótesis, con tal de completar la visión que nos aportan las reglas.

2.1. Durante la formación

En este primer apartado se han considerado las variables de realización de prácticas fueran estas contempladas en el plan de estudios o no; el hecho de haber trabajado de forma remunerada durante la carrera, sin considerarse este trabajo como prácticas; en torno a estos que dicen haber trabajado, una variable sobre si este estaba relacionado con la carrera y si, habiendo continuado en el trabajo 6 meses después de la finalización de la carrera, hubo una mejora de las condiciones de trabajo.

En cuanto a la realización de prácticas, la regla 13 indica una relación entre el hecho de ser mujer y la realización de prácticas.

Figura II-2. Realización de prácticas durante la carrera (base: 29023)



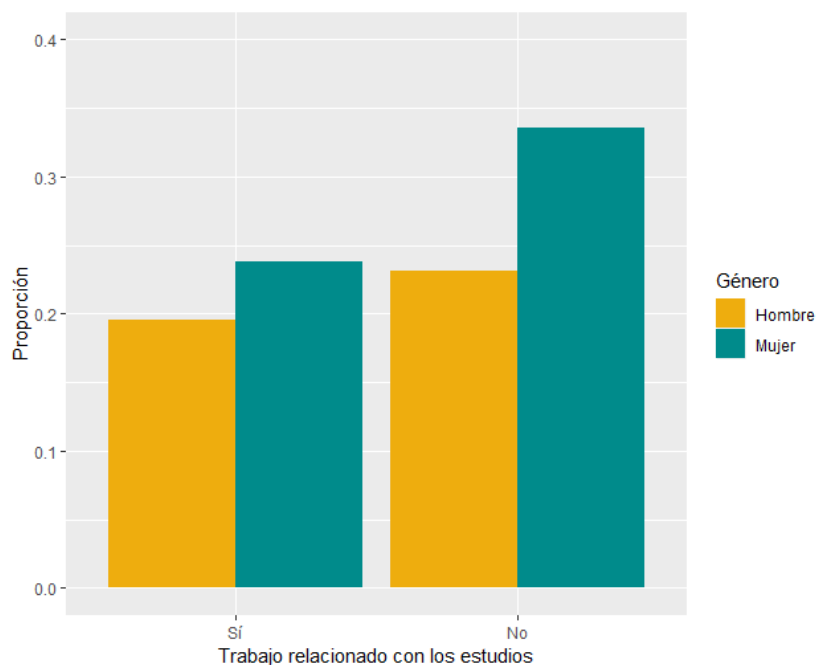
Fuente: Elaboración propia

Con el gráfico se visualiza esta relación, donde ellas serían mayoría entre los que realizan prácticas en empresa durante la realización de la carrera.

Por otro lado, en cuanto a la compaginación de trabajo y estudios, sin ser prácticas, se encuentra relación entre los hombres y el trabajo. Por lo tanto, cuando se habla de trabajo, sin ser prácticas, ellos serían mayoría.

Sin embargo, cuando se analiza si este trabajo está relacionado con los estudios en curso, ellas tienen menor presencia de la esperada en el sí. Es decir, ellos tienen en mayor medida trabajos relacionados con sus estudios.

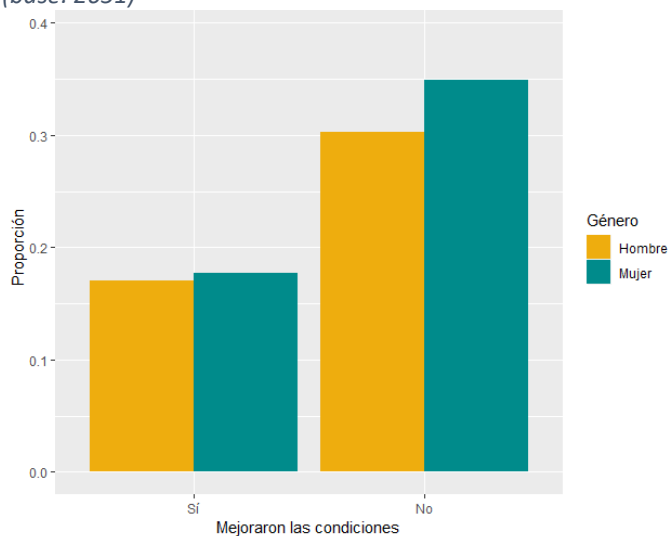
Figura II-3. Trabajo relacionado con la carrera (base: 16930)



Fuente: Elaboración propia

En cuanto a la mejora de condiciones, teniendo en cuenta aquellos y aquellas que han

Figura II-4. Mejora de condiciones al acabar la carrera (base: 2651)



Fuente: Elaboración propia

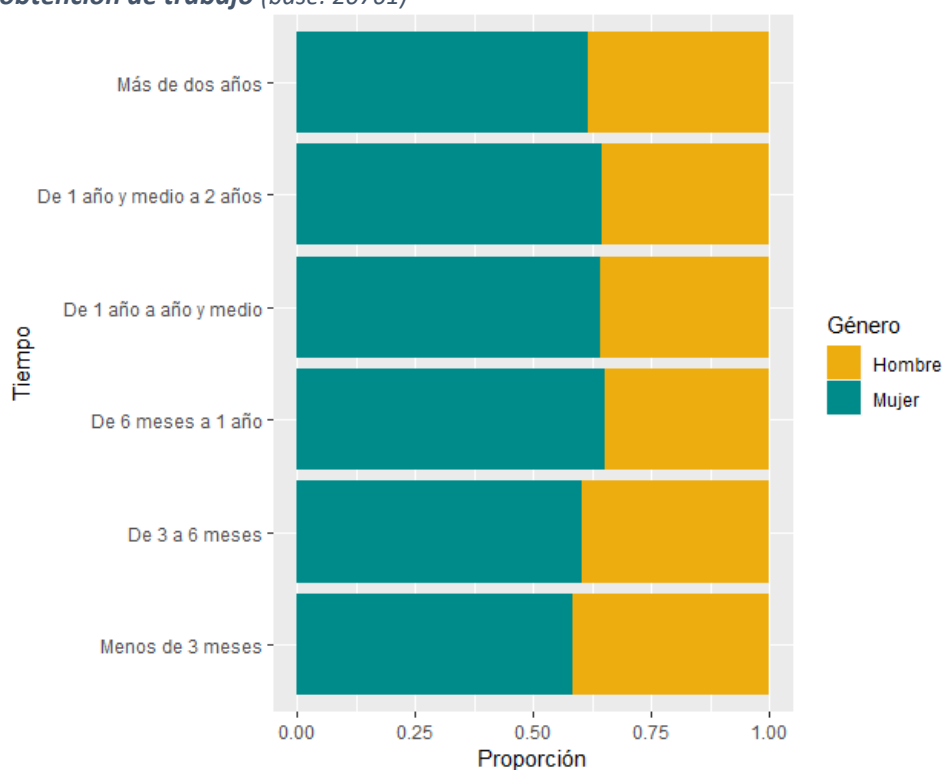
trabajado durante la carrera en un trabajo relacionado con su carrera, se encuentra una mayor presencia de mujeres entre los que dicen no haber tenido una mejora de condiciones una vez finalizados sus estudios.

2.2. Incorporación al mercado laboral

Siguiendo con el análisis de la inserción laboral, se estudian las siguientes variables en referencia al momento de incorporación al mercado de trabajo. Se han tenido en cuenta las variables de tiempo entre finalización de la carrera y trabajar; si se ha rechazado algún empleo desde la obtención de la titulación por considerarlo poco adecuado y, si se ha producido, los motivos de este rechazo; y las maneras de búsqueda del primer empleo.

En cuanto al tiempo transcurrido entre la finalización de la carrera y el momento en el que se empezó a trabajar, se obtiene lo siguiente.

Figura II-5. Tiempo transcurrido entre la finalización de la carrera y la obtención de trabajo (base: 20761)



* se han excluido los individuos que continuaron en el trabajo que tenían durante los estudios

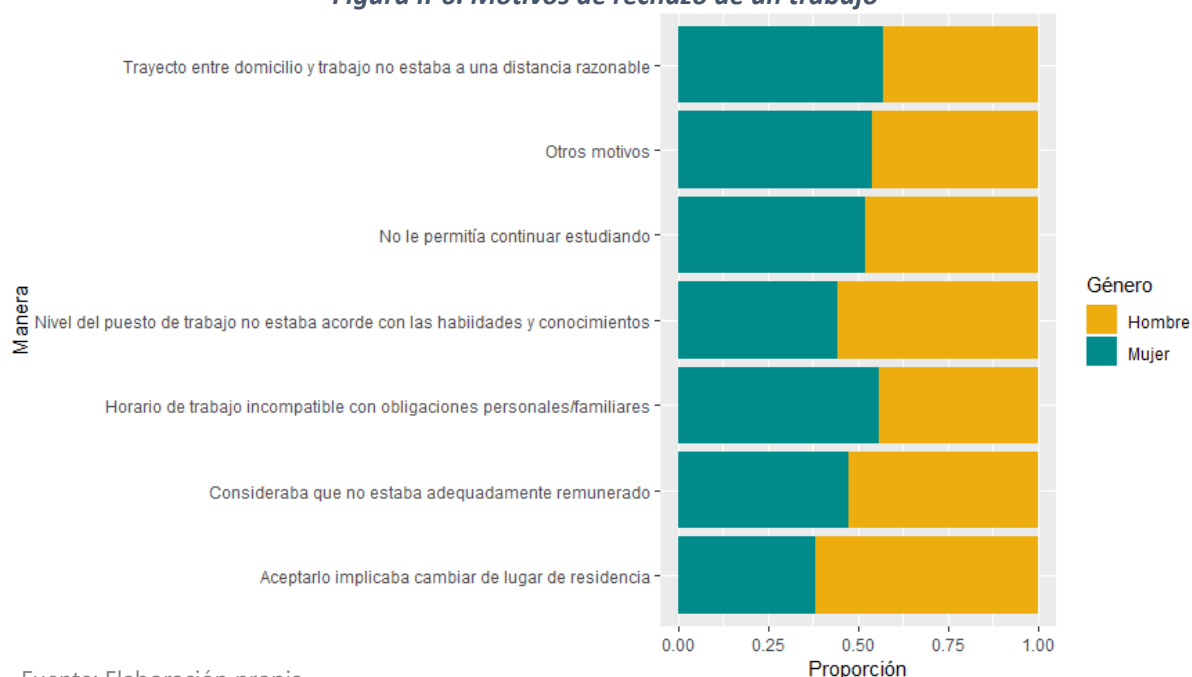
Fuente: Elaboración propia

Se observa una mayor presencia de mujeres en todas las categorías, así como una leve tendencia con presencia creciente de mujeres a medida que aumenta el tiempo. Realizando el test se observa que existen diferencias entre géneros, habiendo una mayor presencia de hombres en las categorías más bajas y mayor presencia de mujeres en las más elevadas.

Volviendo a las reglas, en cuanto al rechazo de un puesto de trabajo por considerarlo poco adecuado, se encuentra asociación entre el ser mujer y el no haber rechazado trabajos desde

la finalización de la carrera. Entrando en los motivos de rechazo, que no aparecen en las reglas de asociación, se observa lo siguiente.

Figura II-6. Motivos de rechazo de un trabajo



Mediante el test de independencia para cada una de las variables según el género, se han encontrado dependencias significativas entre las mujeres y los motivos incompatibilidad horaria y obligaciones personales, trayecto domicilio-trabajo a una distancia no razonable y otros motivos; en referencia a los hombres resultan mayoría en los motivos implicaba cambio de residencia, consideraba que no estaba adecuadamente remunerado y no estaba acorde con las habilidades y conocimientos.

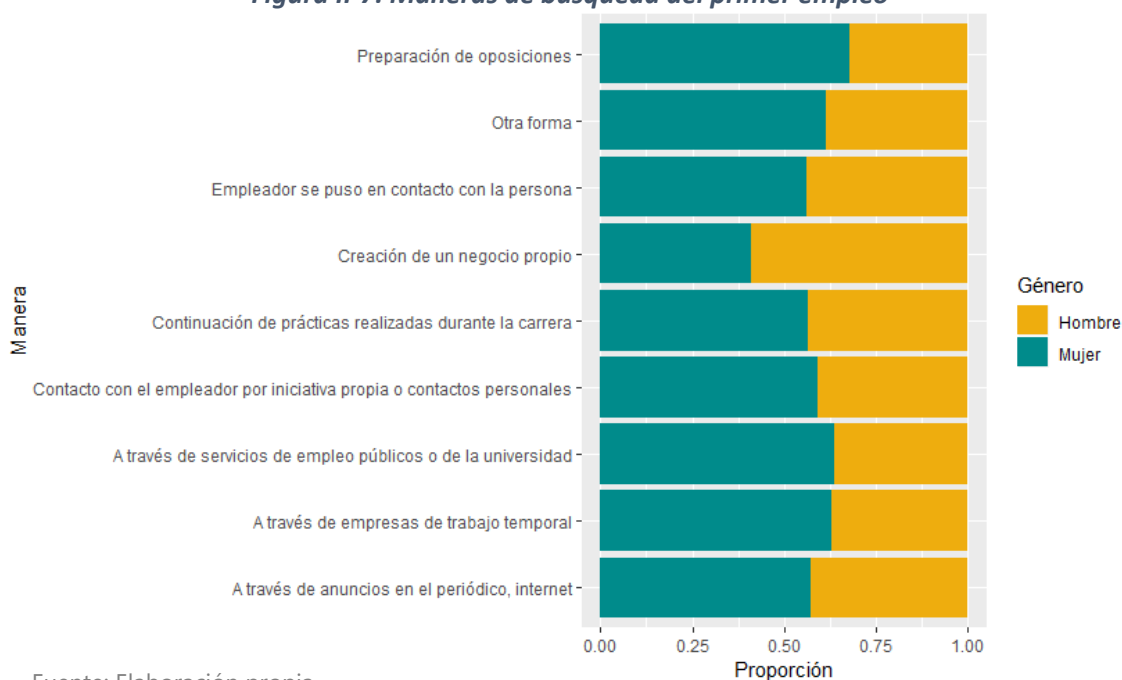
Con estas dependencias, aparecen características históricas diferenciales entre ambos géneros. Ellas hacen mayor alusión a cargas personales o familiares o, en general, motivaciones que no dependen tanto de ellas mismas. Mientras que ellos denotan una actitud más en posesión de sus propias decisiones, como sería el rechazo por la no adecuación de la remuneración o no adecuado a sus capacidades, lo que destaca una actitud con mayor exigencia hacia los puestos de trabajo.

Esto, si bien es cierto, también puede tener una estrecha relación con el ámbito de la profesión. En sectores con menos profesionales y más en auge se puede ser más exigente hacia los trabajos, como podría pasar en el ámbito de las TIC, altamente relacionado con el género masculino. Mientras que en áreas más normalizadas y con mayor disponibilidad de trabajadores, estas exigencias se deberían ver reducidas simplemente por el funcionamiento

del mercado. Este fenómeno podría ocurrir en sectores como la educación, muy relacionado con las mujeres.

De nuevo recurriendo a las reglas obtenidas, observamos la interacción de los géneros con algunas de las maneras de búsqueda del primer empleo de los individuos. Se encuentra una asociación entre el hecho de ser hombre y no haber buscado el primer empleo a través de servicios de empleo públicos o de la universidad, preparando oposiciones o a través de empresas de trabajo temporal. También se observa relación entre el hecho de no haber buscado trabajo contactando con el empleador por iniciativa propia o mediante contactos personales. Así mismo, se obtiene relación entre ser mujer y no haber buscado trabajo a través de anuncios en el periódico o en internet, habiéndose puesto el empleador en contacto con la persona, como continuación de prácticas realizadas durante la carrera o montando un negocio propio.

Figura II-7. Maneras de búsqueda del primer empleo



De todas estas reglas obtenidas, se realiza el test de independencia chi-cuadrado y se obtienen resultados complementarios en casi todos los casos. Se encontraría una mayor presencia, significativa, de mujeres en las variables vinculadas al "no" de los hombres: a través de servicios de empleo públicos o de la universidad, preparando oposiciones o a través de empresas de trabajo temporal. De la misma manera, ellos tendrían presencia significativa en algunas de las variables vinculadas al "no" de ellas: a través de anuncios en el periódico o internet, poniéndose el empleador en contacto con la persona o montando un negocio propio.

Estas maneras de búsqueda de trabajo podrían relacionarse, una vez más, con fenómenos ya conocidos entre hombres y mujeres, como la mayor presencia de mujeres en la educación, muy vinculada a la realización de oposiciones, o bien, el carácter emprendedor de ellos, con mayor presencia en la creación de un negocio propio.

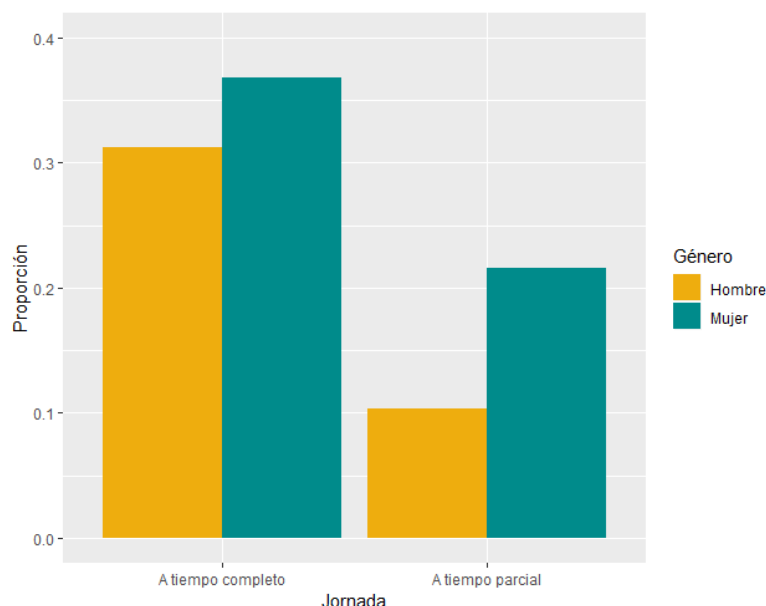
2.3. *Primer empleo*

A continuación, se analizarán las variables en referencia al primer empleo como la situación profesional, la jornada laboral y si se en este trabajo de hacia uso de conocimientos y habilidades adquiridos durante la carrera.

En cuanto a la situación profesional, que hace referencia al tipo de relación contractual con el empleador, se observa relación entre estar asalariado con contrato temporal y ser mujer. Esto denotaría que son ellas que, en su primer empleo, tienen en mayor medida contratos temporales y, con esto, se produciría diferenciación entre géneros, teniendo ellas contratos con “peores características”.

Por otra parte, en cuanto a la jornada laboral, se obtiene una regla que relaciona a los hombres con la jornada a tiempo completo. A continuación, se observa el gráfico de la variable en su

Figura II-8. Tipo de jornada en el primer empleo (base: 18839)

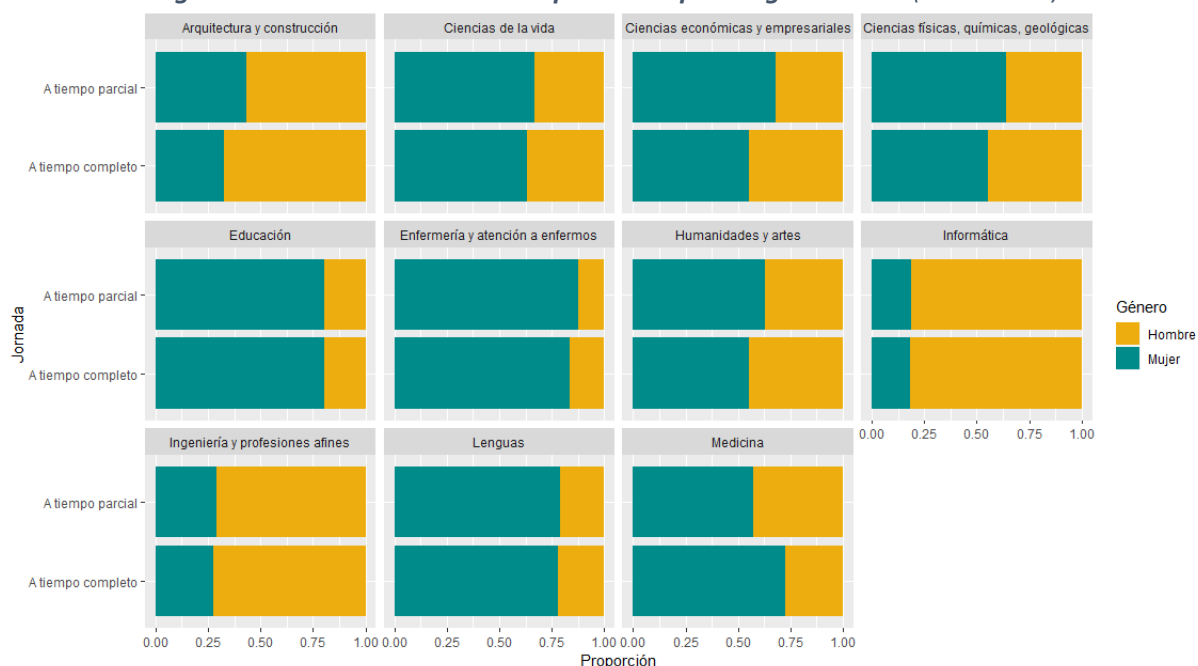


Fuente: Elaboración propia

totalidad, para observar más al detalle este fenómeno.

En este se observa una mayor diferencia entre hombres y mujeres en la categoría de tiempo parcial respecto a la de tiempo completo. Se concluye una mayor presencia significativa de hombres, como ya nos indicaba la regla obtenida, en las jornadas a tiempo completo y una mayor presencia de mujeres en las jornadas a tiempo parcial.

Figura II-9. Jornada laboral en el primer empleo según ámbito (base: 18839)



Fuente: Elaboración propia

Profundizando más en esta variable, se analizará esta según ámbito, con la intención de ver si esto podría incidir en estas diferencias.

Con el gráfico y los tests realizados, se encuentran diferencias significativas en arquitectura y construcción, ciencias económicas y empresariales, ciencias físicas químicas y geológicas, humanidades y artes y en medicina. Estas diferencias son en todos los casos con mayor presencia de hombres en la jornada completa, excepto en medicina, donde serían ellas mayoría en este tipo de jornada.

Tabla II-2. Presencia de géneros en cada ámbito (%) (base: 18839)

Ámbito	Hombres	Mujeres
Arquitectura y construcción	64.68	35.32
Ciencias de la vida	35.61	64.39
Ciencias económicas y empresariales	41.72	58.28
Ciencias físicas, químicas y geológicas	42.04	57.96
Educación	19.73	80.27
Enfermería y at. enfermos	15.53	84.47
Humanidades y artes	41.44	58.56
Informática	81.52	18.48
Ingeniería y afines	72.09	27.91
Lenguas	21.51	78.49
Medicina	28.65	71.35

Para intentar entender un poco mejor este fenómeno, se presenta la tabla de proporciones de hombres y mujeres dentro de cada ámbito.

Exceptuando el ámbito de arquitectura y construcción, donde se encuentra una amplia mayoría de hombres, los demás ámbitos donde se encontraban diferencias en el tipo de jornada entre hombres y mujeres, hay

una leve mayoría de mujeres (60-40). Siendo medicina el único ámbito donde se producen

diferencias a la inversa, siendo los hombres mayoría en la jornada parcial, es uno de los ámbitos con mayor presencia femenina. Estos datos no conducen a una explicación clara de lo sucedido, dejando ver que el ámbito no sería el motivo de estas diferencias. Este fenómeno quedaría sujeto a un estudio más en profundidad.

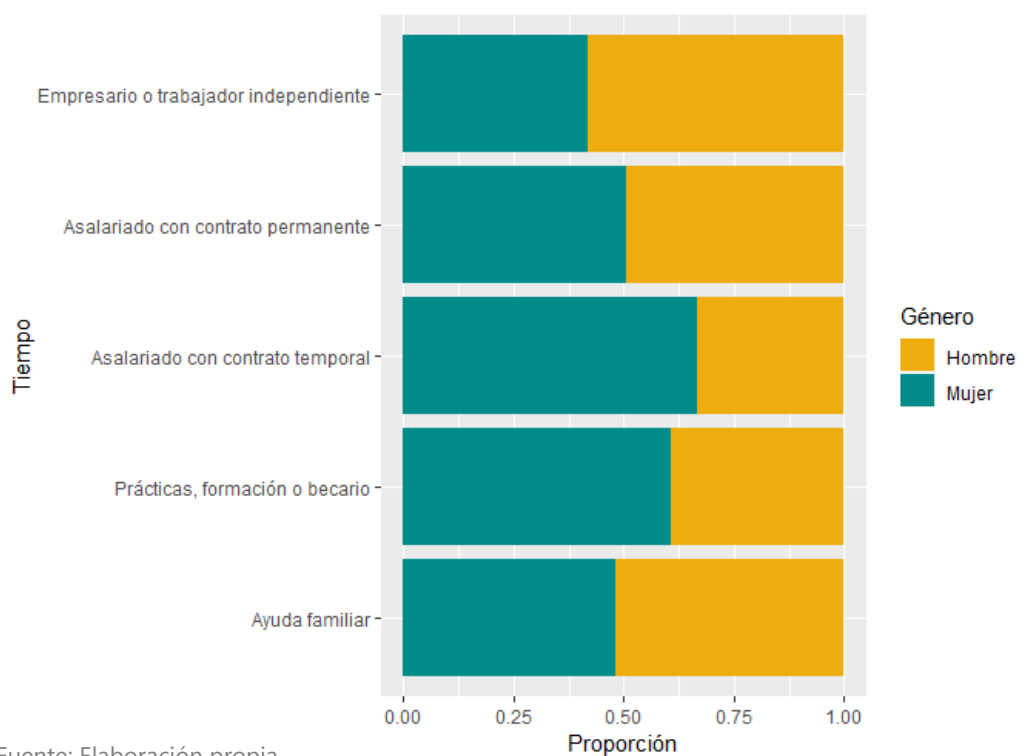
Para finalizar este apartado, se observa en las reglas la aparición de la variable de uso de conocimientos y habilidades adquiridos durante la carrera en el primer empleo. En estas aparece la relación entre el hecho de ser hombre y el uso de estos conocimientos y habilidades en el primer empleo, denotando ellos, una mayor adecuación a sus puestos de trabajo.

2.4. Empleo actual

A continuación, finalizando con el análisis de variables sobre inserción laboral, se estudiarán diversas características sobre el empleo en el momento de la encuesta de los entrevistados. Las variables contempladas en este apartado son la situación profesional del empleo, el tipo de jornada, el número de trabajadores de la empresa y los factores que han influido para conseguir el empleo. También se estudiará la variable que recoge el tiempo trabajado desde la finalización de la carrera.

En cuanto a la situación profesional, se observa lo siguiente.

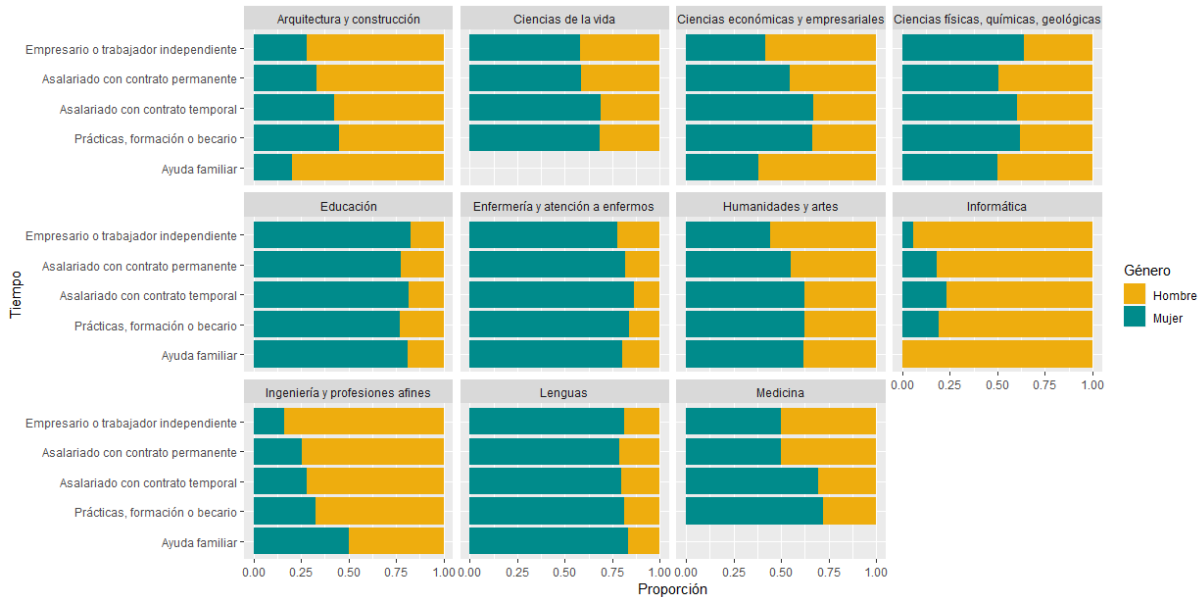
Figura II-10. Situación profesional en el empleo actual (base: 15044)



Ellas tienen, en mayor medida, contratos temporales o de prácticas, formación o becarias, mientras que ellos obtienen más contratos permanentes, son empresarios o trabajadores independiente o ayudan a familiares. De nuevo, se sigue observando unas “peores” característica en torno a los contratos laborales de ellas, mientras que ellos tienen o contratos más “estables” o ese carácter emprendedor, que abunda más entre hombres.

Este hecho sería interesante observarlo también por ámbitos.

Figura II-11. Situación profesional en el empleo actual según ámbito (base: 15044)



Fuente: Elaboración propia

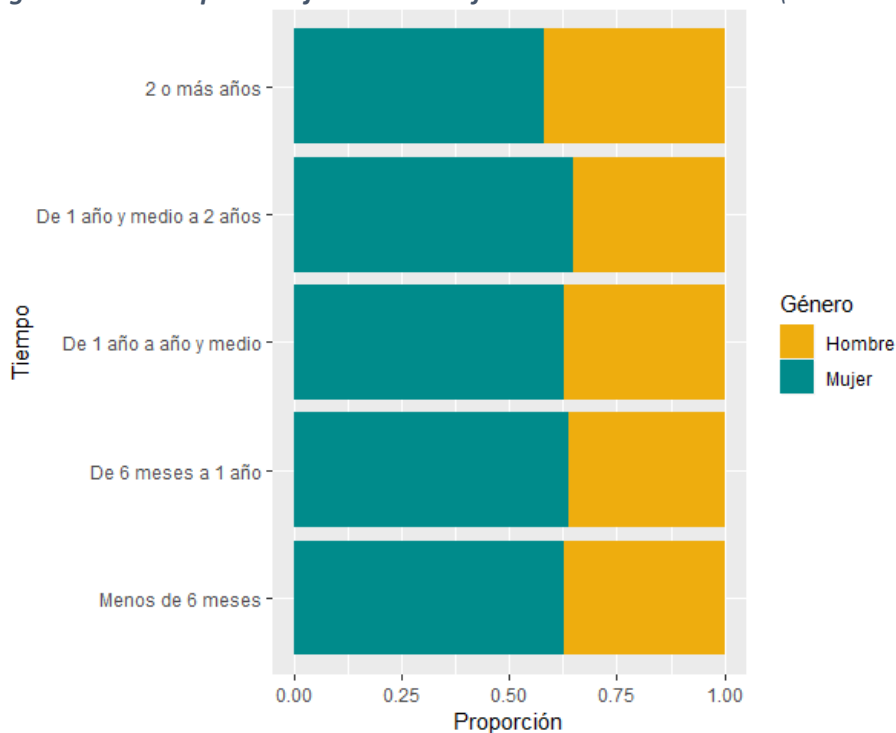
En todos los casos que se encuentran diferencias, estas son favorables a los hombres (mayoría con contrato permanente o empresarios) y con peores condiciones para las mujeres (contrato temporal o prácticas). Estas se encuentran en los ámbitos de arquitectura y construcción, ciencias económicas y empresariales, humanidades y artes, informática y en ingeniería y profesiones afines. De nuevo se encuentra mayoría de mujeres con peores condiciones profesionales tanto en ámbitos muy masculinizados como en sectores donde existe una leve mayoría femenina. Con esto no se podría explicar ningún patrón de diferencias en relación con el ámbito.

Volviendo a los resultados obtenidos mediante reglas de asociación, ellos tienen en mayor medida jornadas a tiempo completo, lo que seguiría afirmando lo comentado en el punto anterior en torno a la situación profesional de ellas y ellos.

Siguiendo con la dimensión de la empresa, no se encontrarían diferencias entre hombres y mujeres. Ellos y ellas se encuentran presentes de la misma manera en todo tipo de empresas, en lo que a número de trabajadores se refiere.

Por otro lado, se estudia el tiempo trabajado desde la obtención del título, con la intención de observar si mujeres y hombres tienen tendencias diferentes en el inicio de sus carreras profesionales.

Figura II-12. Tiempo trabajado desde la finalización de la carrera (base: 18929)

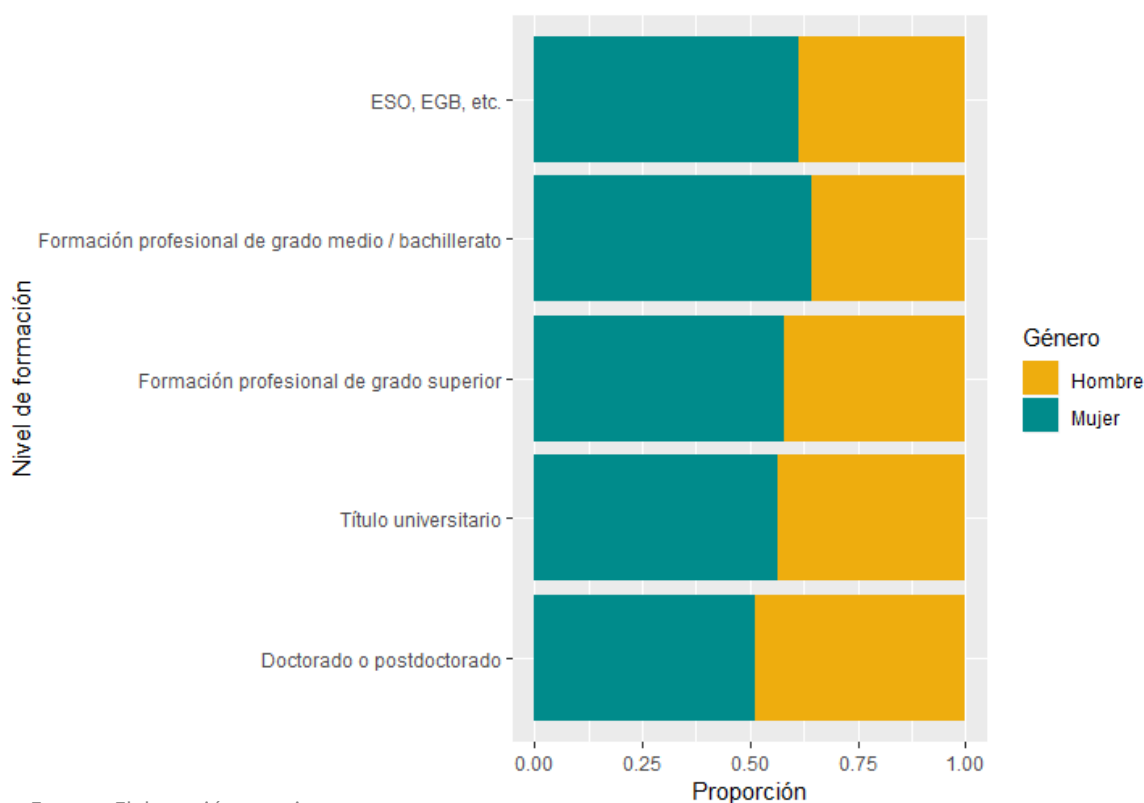


Fuente: Elaboración propia

En torno al tiempo, se encuentran diferencias significativas entre hombres y mujeres, habiendo una mayor presencia de mujeres en todas las categorías excepto en la mayor (2 años o más). Desde la etapa inicial de la carrera profesional de los estudiantes universitarios, ellos trabajarían más tiempo, posiblemente gracias a un inicio más temprano.

Finalmente, se analiza la percepción que tienen hombres y mujeres en cuanto al nivel de formación más apropiado para realizar el trabajo actual, teniendo en cuenta que todos los individuos estudiados son titulados universitarios.

Figura II-13. Percepción del nivel de formación más adecuado para realizar el trabajo actual (base: 14570)



En esta variable se encuentra una mayor presencia de hombres en las categorías superiores "título universitario" y "doctorado o postdoctorado", viendo en estos una buena correspondencia entre sus estudios y el trabajo realizado. Por otro lado, se encuentran ellas sobrerrepresentadas en las categorías inferiores de "formación de grado superior / medio /bachillerato" o "ESO, EGB, etc.", recogiendo como ellas tienen una mala asociación entre estudios y trabajo, puesto que realizan tareas "por debajo de sus posibilidades". Mientras que ellos tendrían un buen posicionamiento, ellas se encontrarían en puestos de trabajo menos adecuados a sus estudios.

A continuación, se presentan las principales conclusiones extraídas del segundo capítulo. Son las mujeres quienes más realizan prácticas durante la carrera, pero son ellos los que más compaginan trabajo y estudios. Entrando en el detalle de estos trabajos se observaba que, cuando hombres y mujeres trabajan durante la carrera, ellos lo hacen en mayor medida en empleos relacionados con sus estudios; también en torno a esto se obtenía que, si estos continuaban en dichos trabajos una vez acabada la carrera, ellos veían mejoradas, en mayor medida, sus condiciones laborales.

Una vez acabada la carrera, se observa como ellas son mayoría entre quienes se demoran más en empezar a trabajar. Sobre el rechazo de puestos de trabajo por considerarlos no adecuados, los hombres presentan una mayor notoriedad en el sí, siendo las mujeres quienes menos rechazan puestos de trabajo; y cuando se producen estos rechazos, se encuentran diferencias claras entre hombres y mujeres sobre los motivos. Ellos denotan motivaciones más relacionadas con decisiones propias, mientras que ellas alegan motivos más externos, no dependiendo de la misma persona.

En cuanto a las condiciones de los puestos de trabajo iniciales, también se encuentran ya diferencias según el género. En torno a las condiciones de los hombres se percibe mayor presencia de contratos permanentes y a jornada completa, mientras que a las mujeres se las asocia más con contratos temporales o en prácticas y a media jornada. Pese a analizar estas diferencias según ámbito, no se ha encontrado ningún claro factor explicativo, produciéndose diferencias tanto en ámbitos más masculinizados como en aquellos con mayor presencia femenina. Finalmente, se percibía una adecuación desigual entre hombres y mujeres a sus puestos de trabajo, ellos tendrían una mayor correspondencia entre trabajo y estudios, mientras que ellas dicen en mayor medida estar en puestos correspondientes a menores niveles formativos.

Una vez finalizados los capítulos 1 y 2, habiendo sido explorados los ámbitos de educación y de inserción laboral, se presentarán, según los resultados obtenidos, los principales puntos donde se han detectado *brechas* educacionales y de inserción laboral entre hombres y mujeres.

En la etapa educativa inicial, se percibían diferenciaciones de niveles, siendo las mujeres las que actualmente se encuentran en mayor medida en los altos niveles de estudio. También se detectaban las ya conocidas *brechas temáticas*, donde siguen existiendo claros ámbitos de hombres y de mujeres. Estos aspectos ya sentarían diferencias iniciales entre hombres y mujeres.

Avanzando en la carrera profesional de las personas, se encontraría el momento de la inserción laboral, donde se percibe, en general, un mejor posicionamiento de ellos y no tan bueno para ellas. Sin ser esto un problema extremo, se encontraba como ellas tenían condiciones más de tipo temporal y ellos condiciones más estables.

También en cuanto a un aspecto más progresivo durante la vida laboral como sería la formación continuada, se encontraban diferencias cruciales. Ellos son mayoría entre los que realizan formación en sus puestos de trabajo, aunque son ellas las que participan en más

actividades educativas no formales, fuera del entorno laboral. Con esto, las empresas pagan actividades formativas, en mayor medida, a hombres que a mujeres.

Relacionando todo esto con la brecha salarial, se plantean las siguientes ideas. En cuanto a educación formal se refiere, se encuentra una *brecha educativa positiva*⁹ hacia las mujeres, siendo ellas las que más tienen altos niveles de estudios, por lo tanto, no habría una relación aparente que explicase la diferencia de salario por la falta de estudios formales y reglados de las mujeres. Saliendo de lo que se considera educación formal y entrando en lo que comúnmente se denomina "formación en la empresa" sí que se encontraría una explicación que contribuiría al entendimiento de la brecha salarial. Ellos serían los beneficiarios en este caso de una *brecha formativa* en cuanto a formación continuada, siendo los hombres quienes más participan en estas actividades educativas proporcionadas u organizadas por la empresa. Esto conduciría a lo largo del tiempo a una mejor preparación de los hombres, un mejor "currículum" y, en definitiva, un mejor perfil profesional que radicaría en una mejor retribución monetaria.

El factor de la participación de hombres y mujeres en los diferentes ámbitos y sectores profesionales es donde se establece, según los resultados obtenidos, una clara *brecha temática*. Siguen existiendo claras áreas de mujeres y hombres. Una vez finalizado el primer capítulo se establecía una hipótesis en torno a la posibilidad de explicación de muchas diferencias debido a esta distribución tan marcada de hombres y mujeres en determinados sectores. A lo largo del capítulo 2, esto no ha podido ser probado, sin ser encontradas asociaciones entre sectores-diferencias y sin discernir patrones claros en torno a esto. Aquí se establece una clara futura línea de investigación.

Finalmente, en cuanto a la inserción laboral, también se podría establecer una *brecha contractual* en los inicios profesionales de mujeres y hombres. Entre las condiciones de ellos y ellas se encuentran diferencias, siendo el género masculino el que presenta un mayor posicionamiento en las empresas, con mejores condiciones y mayor adecuación.

⁹ Se denomina positiva, puesto que se parte de la premisa que son las mujeres las que sufren, en mayor medida, las consecuencias de la brecha salarial y, cuando las diferencias se producen a la inversa, esto se denomina positivo, en cuanto al cambio de objeto que se ve beneficiado de dicha diferencia

Capítulo III. Análisis de concienciación

Finalmente, en este capítulo final, se explorará una tercera vía, con la intención de utilizar métodos estadísticos más novedosos y aprovechar el llamado *Big Data*.

En un primer momento se intentó explorar la API de LinkedIn. Puesto que la red social profesional actual por excelencia es LinkedIn, hubiera sido muy interesante estudiar y analizar mediante esta las actitudes de las personas durante su carrera profesional. Debido a la sensibilidad de los datos, no es fácil a nivel usuario acceder a ellos de manera gratuita, y dadas las limitaciones que se tienen al realizar un trabajo final de grado en solitario, finalmente no fue posible obtener datos útiles para el objeto de estudio, que sería tener variables concretas de interacción entre usuarios, empleos y demás según su género. Con todo esto, se plantea una futura línea de investigación, pudiendo ser una herramienta muy útil para el estudio y la detección de posibles brechas o diferenciaciones entre hombres y mujeres en su carrera profesional.

Ante este impedimento, se plantean dos fuentes de datos alternativas. La red social Twitter y el buscador Google, usando estos como referentes de redes sociales y buscadores. Con dificultades similares a las anteriores, puesto que no se puede obtener de manera fácil y gratuita acceso a muchos datos, se explorarán estos dos medios, no con la intención de extraer claros resultados, sino con el propósito de tener una visión global y diferente sobre el contexto estudiado.

3.1. Twitter

Así pues, se ha procedido a analizar mediante Twitter, la opinión o interacción de los usuarios durante 14 días¹⁰ con los siguientes hashtags: #BrechaSalarial, #InsercionLaboral. También se ha recogido la interacción con el hashtag #Igualdad durante 7 días.

En el caso de los dos primeros, se recogieron datos durante dos semanas consecutivas, para obtener un mayor número de tweets, en la medida de lo posible, y poder llegar a un mejor análisis. Por otro lado, en cuanto al uso de #Igualdad, puesto que es un hashtag más utilizado, se ha considerado suficiente el análisis de los tweets durante una semana.

A continuación, mediante la nube de palabras sobre los hashtags estudiados, se analizarán las palabras más usadas.

¹⁰ El perfil de desarrollador gratuito de Twitter solo permite acceder a los tweets publicados en los últimos 7 días

En la semana de referencia, se realizaron 7301 tweets con la aparición de la etiqueta #Igualdad. En estos se utilizaron un total 86133 palabras.

[illegible]

Como era de esperar, este es el termino en torno al cual se encuentra una mayor interacción. Además, se encuentra una estrecha relación entre las palabras que aparecen conjuntamente al término y la temática, es decir, se puede percibir como esta palabra tiene la intención bien arraigada entre los usuarios. Se observan palabras como "feminismo", "mujeres" o "heteropatriarcado", entre otras. Estas demuestran, como ya se ha comentado, una buena atracción temática de este hashtag y, por lo tanto, concienciación en las redes sociales en cuanto a la igualdad, puesto que cuando se utiliza este hashatg, se hace con la intención de hablar de este tema en concreto y ese no sería confundido o aparecido en otro tipo de comentarios.

Figura III-2. Nube de las palabras más usadas en los tweets con hashtag #BrechaSalarial



Finalmente, en torno a la temática de la Inserción laboral se han encontrado en las 2 semanas estudiadas un total de 132 tweets, conteniendo 1561 palabras. De nuevo se observa una clara disminución de uso. En este caso no se podría relacionar con una no-concienciación en torno

Figura III-3. Nube de las palabras más usadas en los tweets con hashtag #InsercionLaboral



Con esto, se determinaría que no existe una asociación en la sociedad de este término con ninguna problemática en relación a diferencias de género.

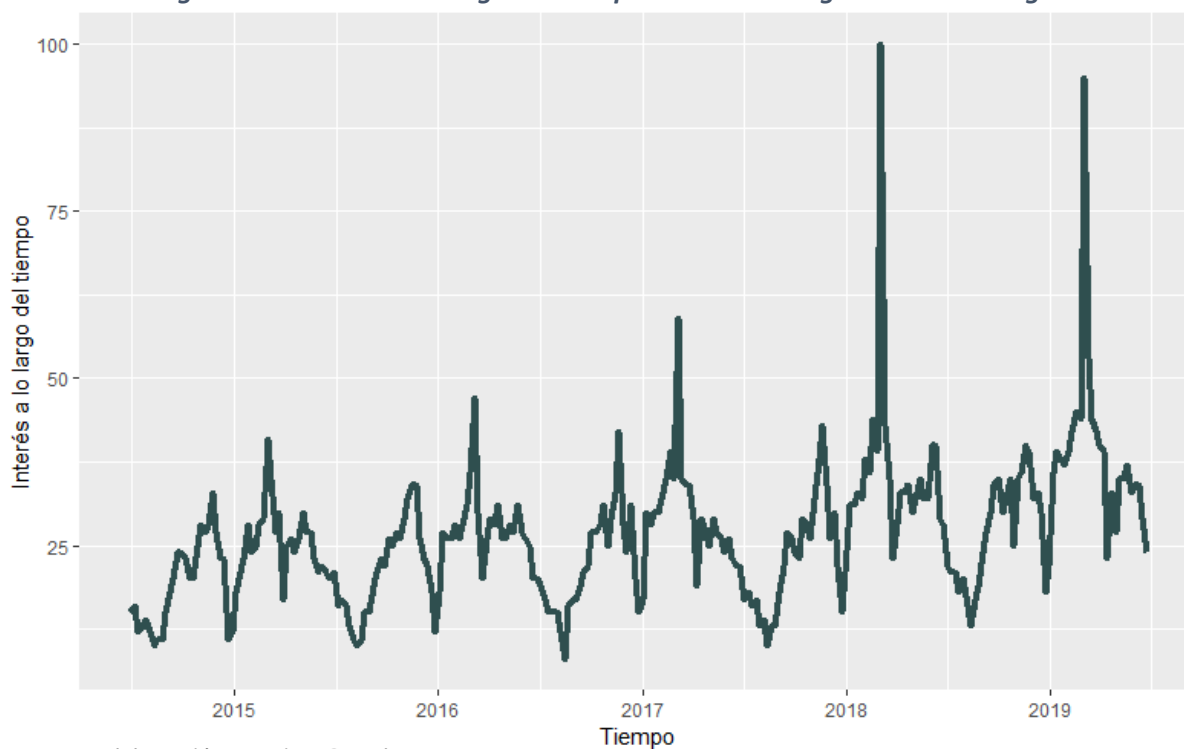
De Google, se ha extraído la serie temporal del interés de búsqueda en relación con el valor máximo de los datos en el periodo en cuestión. El 100 indicaría la popularidad máxima en el momento de más búsquedas, mientras que 50 indicaría que un término es la mitad de popular en relación con el valor máximo.

68

han recogido por separado, demostrándose el interés de búsqueda de cada término sobre sí mismo, produciéndose así el mismo efecto que observándose los datos de búsquedas absolutos. También en la parte final se han juntado los tres términos, donde sí que se mide el interés sobre una misma base, con la intención de comparar las series temporales.

En primer lugar, se analiza la serie temporal de la búsqueda en Google de la palabra "Igualdad". Ya en Twitter se ha podido observar como este era el término con más interacciones por parte de los usuarios y, a continuación, se buscaría recoger como esta interacción ha evolucionado en los últimos años.

Figura III-4. Interés a lo largo del tiempo del término "Igualdad" en Google



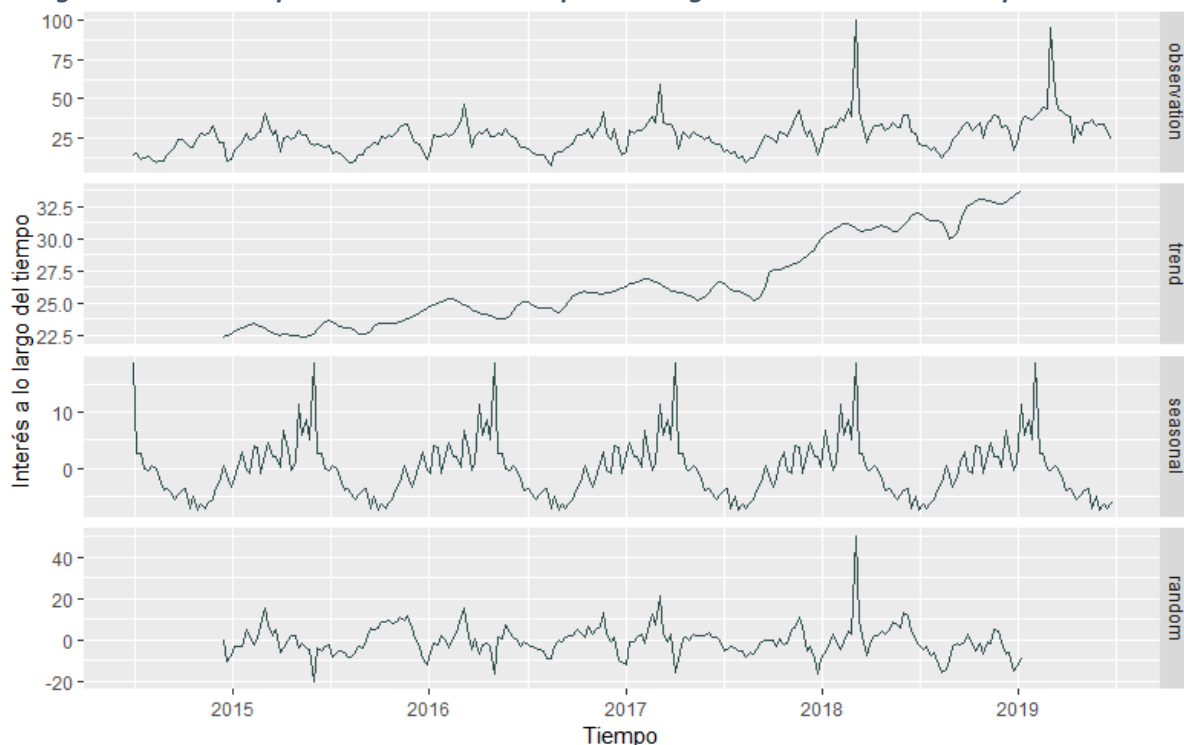
Fuente: Elaboración propia / Google

En esta serie temporal se perciben diversos aspectos. El primero sería un posible carácter estacional y una aparente tendencia, muy influenciada por picos concretos muy marcados, que crecen en diferencia a lo largo de los años. Estos "picos" coinciden con la semana del 8 de marzo, día de la mujer trabajadora, que se ve como con los años va influyendo más, teniendo los años 2018 y 2019 una gran influencia sobre la búsqueda de este término. Esto se relaciona con una creciente concienciación de la sociedad española sobre el objetivo, todavía pendiente,

de la igualdad entre hombres y mujeres, a lo que contribuye el día de la mujer, día que sociedad y medios aprovechan para hacer eco de este movimiento social.

Para comprobar si esta serie tiene tendencia y estacionalidad se presenta el siguiente gráfico.

Figura III-5. Descomposición de la serie temporal de “igualdad” mediante un esquema aditivo



Fuente: Elaboración propia / Google

Mediante la descomposición de la serie siguiendo un esquema aditivo, queda muy clara la componente de tendencia, con un crecimiento a lo largo de los años. Mediante el test de

Ilustración III-1. Resultado del test de Spearman (R)

```
Spearman's rank correlation rho
data: ts_igualdad and time(ts_igualdad)
S = 1553839, p-value = 3.874e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.4756249
```

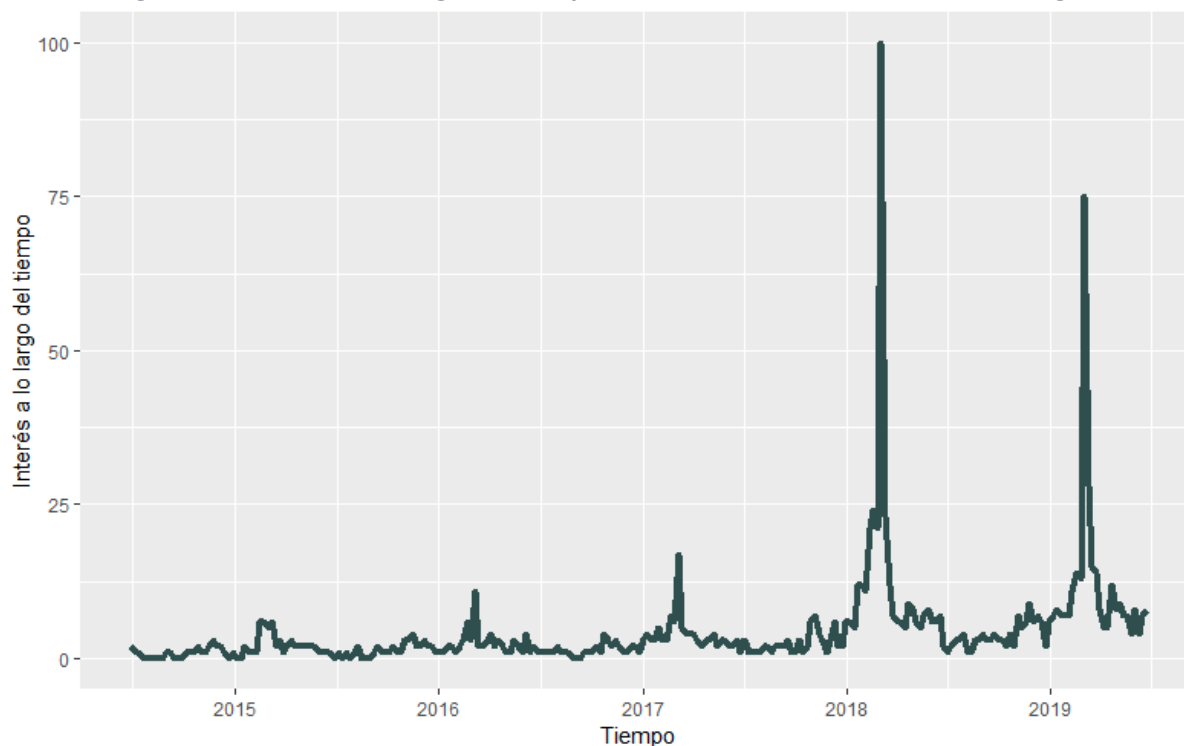
Spearman ya comentado en la metodología, que resulta significativo, se determina que la búsqueda del término igualdad ha crecido en los últimos 5 años y, por lo tanto, a día de hoy hay una mayor concienciación entorno a

esto.

También se observa con la descomposición un claro factor estacional, en relación, como ya se comentaba, al día 8 de marzo. Con esto, se podría afirmar que existe también un componente estacional que sucede en la búsqueda del término “igualdad”. Esto, también se podría extrapolar a la concienciación en general, existiendo “concienciación estacional” en cuanto a la igualdad, en torno al día de la mujer.

A continuación, se analiza de la misma manera el interés temporal en relación con la brecha salarial.

Figura III-6. Interés a lo largo del tiempo del término “brecha salarial” en Google

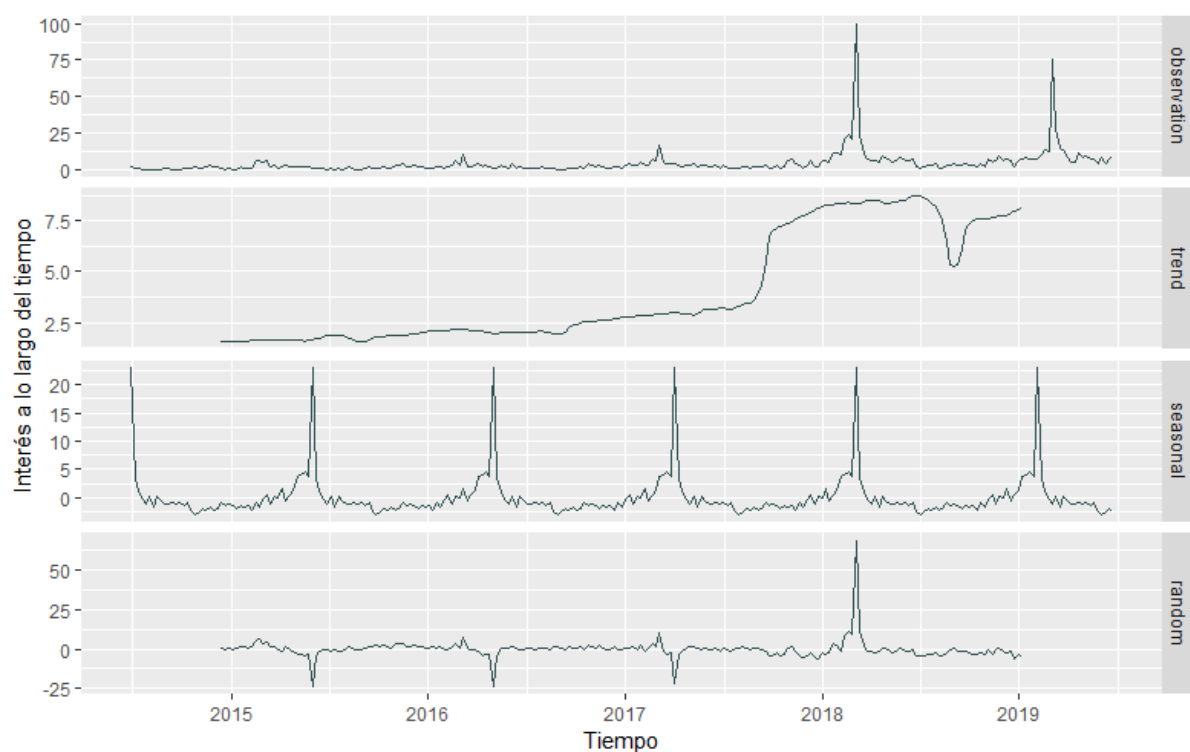


Fuente: Elaboración propia / Google

En esta serie temporal se pueden ver de manera clara los mismos repuntes que en caso de la igualdad, en torno al 8 de marzo. A medida que han aumentado los años se ve cómo ha ido apareciendo este incremento en dicha semana, habiendo un crecimiento muy grande en los últimos años, donde el interés en torno a la brecha salarial se ha disparado.

Para estudiar los componentes estacionales y de tendencia, se analiza el grafico descompuesto.

Figura III-7. Descomposición de la serie temporal de “brecha salarial” mediante un esquema aditivo



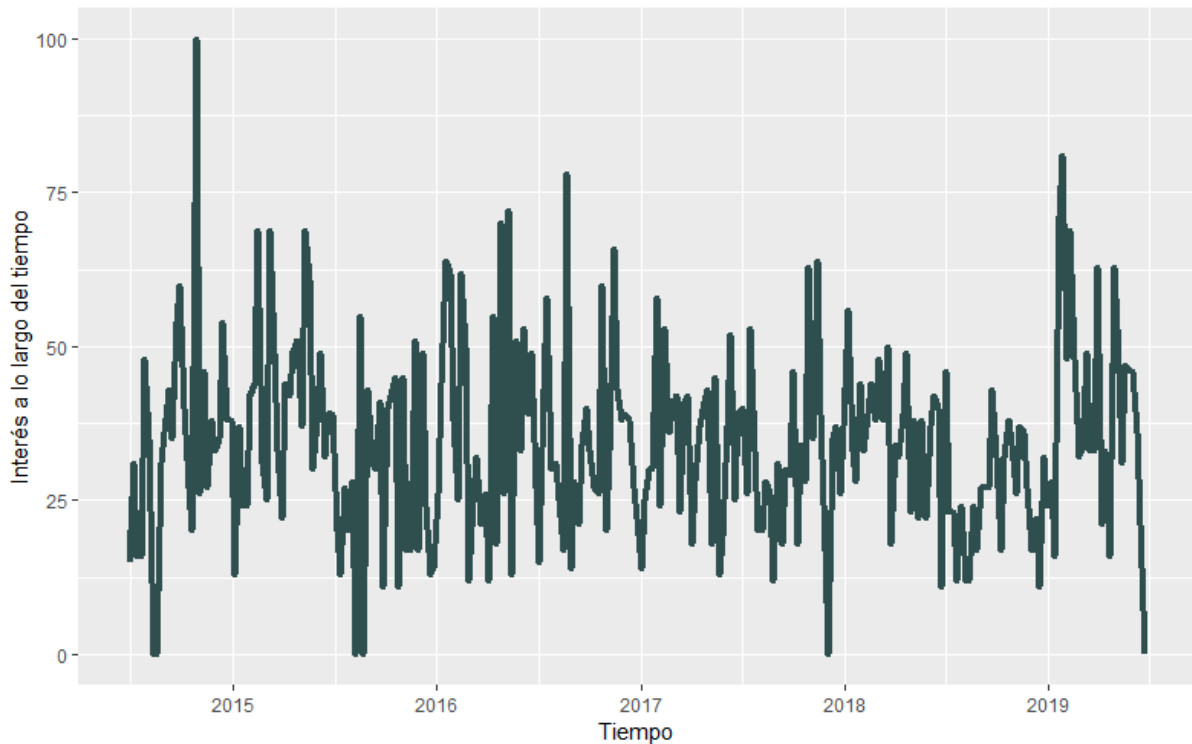
Fuente: Elaboración propia / Google

Aquí se puede ver tendencia creciente, aunque con un comportamiento extraño, debido a esa gran influencia de las búsquedas en los últimos dos años por el día de la mujer. La tendencia quedaría comprobada con el test que resulta significativo, con lo que también habría una concienciación creciente en cuanto a la brecha salarial, también ligada con la concienciación de la igualdad.

También aquí se observa un claro componente estacional. Esto coincidiría con el comportamiento observado con la igualdad, aunque con la igualdad haya sido un proceso más gradual y con la brecha salarial haya surgido más de manera repentina en el espacio temporal estudiado.

Vistas las series temporales de igualdad y brecha salarial, se repetirá el mismo proceso con la inserción laboral.

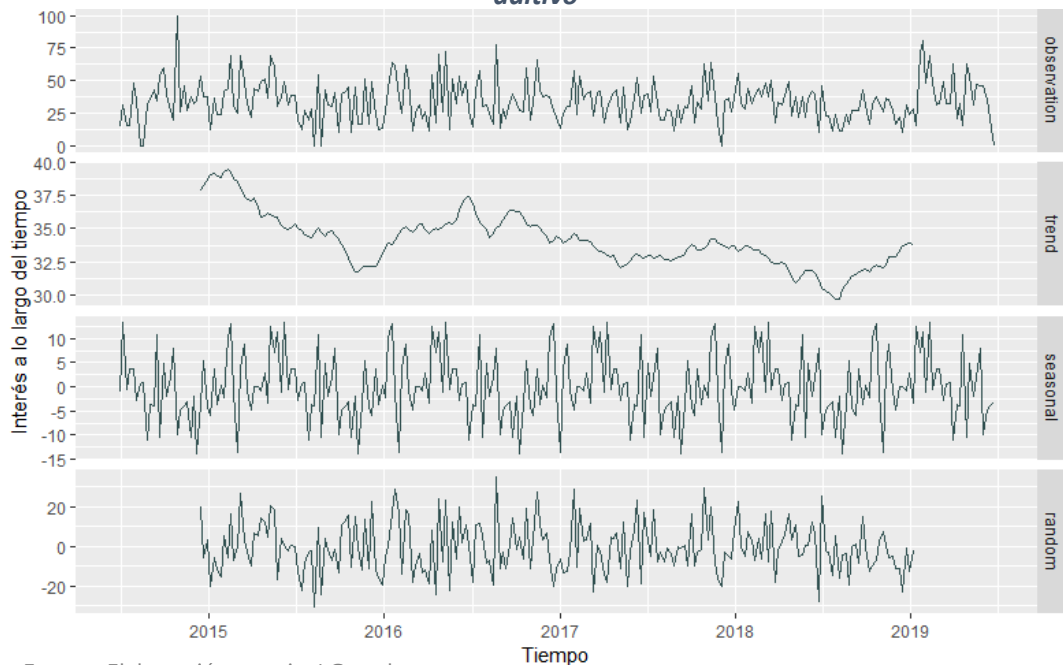
Figura III-8. Interés a lo largo del tiempo del término “inserción laboral” en Google



Fuente: Elaboración propia / Google

En esta serie temporal, no se observa, fenómenos, viéndose casi como un proceso aleatorio. Para determinarlo de manera afirmativa, se analizarán los datos descompuestos.

Figura III-9. Descomposición de la serie temporal de “inserción laboral” con un esquema aditivo

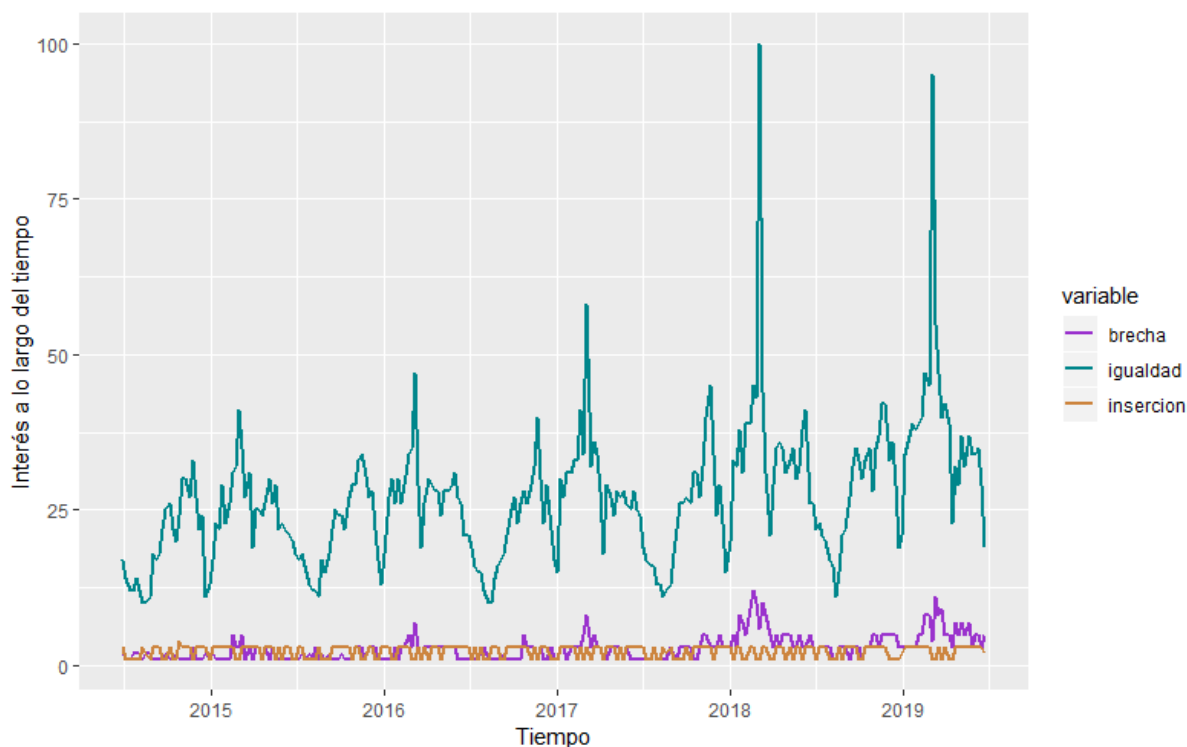


Fuente: Elaboración propia / Google

En estos tampoco se observarían factores de tendencia ni estacionalidad. Este concepto rompería los comportamientos observados en torno a la igualdad y la brecha salarial, denotando una no-concienciación de la sociedad sobre las diferencias en inserción laboral entre hombres y mujeres. Es cierto que esto era de esperar, puesto que la inserción laboral es un hecho mucho más concreto y focalizado, y tampoco está muy extendido entre las personas que exista una problemática en torno a este tema, como ocurre con la brecha salarial, estudiada, difundida y conocida por gran parte de la sociedad actual. También sobre en este término, habría confusión entre diferentes campos que podrían estar implicados, no solo en cuanto al género, sino sobre otras problemáticas del mercado laboral.

Finalmente, se presentan las tres series temporales conjuntas, en base a las búsquedas de la igualdad, por ser aquella con más interacciones en nombres absolutos.

Figura III-10. Interés a lo largo del tiempo de “igualdad”, “brecha salarial” y “inserción laboral”



Fuente: Elaboración propia / Google

En este se recogen los hechos que se han ido comentando, se observa la tendencia creciente y el carácter estacional del término igualdad, establecido en cuanto a concienciación, y a medida que aumentan los años se observan estos comportamientos también en torno a la brecha salarial, creciente y con puntos estacionales claros por el día de la mujer. En cuanto a la inserción laboral, también vinculada con otras problemáticas, no se observa ningún

comportamiento que, a día de hoy, denote algún cambio en cuanto a la concienciación sobre esto.

Para finalizar este bloque, se destacarán las ideas claras extraídas mediante el análisis de Twitter y Google.

Aunque últimamente está más extendido el uso de Twitter para analizar sus datos, en este trabajo habría que destacar la aportación positiva de los datos de Google mediante Google Trends, una herramienta quizá no tan extendida, pero muy útil a la hora de proporcionar una visión general y clara, teniendo en cuenta a una mayor parte de la población respecto Twitter.

Es por eso, que Twitter ha aportado una visión exploratoria muy concreta en cuanto a palabras usadas y interacción en Twitter, pero es gracias a Google que se han podido extraer dos ideas claras en cuanto a la concienciación en internet. Se observa una tendencia creciente en cuanto a la búsqueda de términos concretos como son Igualdad y Brecha Salarial, y también estacionalidad centrada en el día de la mujer. Esto denota que está habiendo un aumento de la concienciación sobre estos temas y que, gracias a la difusión en general, y en torno al día de la mujer en particular, se producen efectos directos sobre ese conocimiento o intención de conocimiento. Conociendo esto, es importante seguir trabajando desde todos los ámbitos posibles por el estudio, la información y la difusión de las áreas donde se producen diferencias de géneros y, contribuir así a la igualdad entre hombres y mujeres.

Conclusiones

Finalizando el trabajo realizado, a continuación se presentan los principales resultados, dando respuesta a las hipótesis, cuestiones y objetivos planteados.

Se han detectado diversos factores en distintos momentos de la carrera profesional de las personas donde hombres y mujeres no son iguales. Estos estudian diferente, tienen diferencias a la hora de inserirse en el mundo laboral y no tienen las mismas condiciones laborales en el momento inicial de sus carreras profesionales, incluso demuestran actitudes diferentes ante educación y formación.

Con este trabajo, se concluye la existencia de una *brecha educativa*, siendo las mujeres las que mejor nivel de estudios obtienen, dentro de la cual también se establece una *brecha temática* muy marcada, con un largo recorrido histórico y todavía presente entre los estudiantes actuales.

También existe una *brecha formativa* en cuanto a la formación continuada, siendo los hombres lo más beneficiados a lo largo de sus carreras por parte de las empresas.

De la misma manera, se establece una tercera *brecha contractual* existiendo diferencias en las condiciones laborales de hombres y mujeres, desde el primer momento en que estos pasan a formar parte del mundo laboral.

Con estas conclusiones y volviendo a la idea principal de conocer mejor el contexto profesional de las personas, que se conoce que lleva a la brecha salarial, se destaca lo siguiente.

Por un lado, si solo se observara el entorno educativo, tendría lógica que, en la sociedad en general, las mujeres se encontraran mejor posicionadas en el mundo laboral.

Si se avanza, observando la formación continuada o la inserción al mundo laboral, ya se empiezan a ver factores donde ellos toman una mejor posición en condiciones laborales, con diferentes aspectos que intervienen: formación en la empresa, condiciones contractuales o adecuación al puesto de trabajo. Es en estos factores donde se inician las diferencias negativas hacia las mujeres, y que radicarían en lo que ya se conoce, que es una diferencia en las retribuciones de hombres y mujeres.

Finalmente, se ha visto la tendencia creciente de la concienciación en los últimos años en torno a la igualdad o la brecha salarial. También sale a relucir el componente estacional, que destaca el repunte de interacción que se produce alrededor del 8 de marzo, día de la mujer trabajadora, muy útil, en este caso, para la difusión de problemas sociales que inciden en las desigualdades entre mujeres y hombres.

Políticas sociales y futuras líneas de investigación

Finalmente, con la intención de aplicar todas las conclusiones extraídas, se propondrán líneas de actuación en torno a las problemáticas o diferenciaciones encontradas, así como futuras líneas de investigación surgidas durante la realización del trabajo. Siendo esta actuación la culminación del trabajo siguiendo el siguiente esquema:



Como es conocido, los estudios estadísticos de los datos son una vía muy positiva y efectiva de detección de problemáticas y establecimiento de propuestas de solución a estas. Estando el trabajo enfocado a un tema sociológico y, por lo tanto, siendo más compleja la definición de unas líneas de actuación claras para la solución de la problemática, se tratarán de definir puntos claros donde se han encontrado diferencias, basados en los resultados obtenidos y, consecuentemente, donde se encontraría una necesidad de actuación.

Como ya se ha comentado, se determina la existencia de una *brecha educativa*, una *brecha temática*, una *brecha formativa* y una *brecha contractual*.

Con esto, se concluye que sería necesario incidir no solo en mejorar las condiciones salariales de las mujeres, hecho importante, sino también en mejorar todos los factores que ocurren antes o durante.

En términos de educación es importante que ellos y ellas tengan las mismas oportunidades, sin darse otros factores que los condicionen en la primera etapa y el punto de partida, en la mayoría de ocasiones, como son los estudios escogidos. También se consideraría importante, vistos los claros sectores masculinizados y feminizados, que los estigmas o prejuicios que llevan a estudiantes a decantarse por un ámbito u otra se erradiquen. En torno a esto no se puede esperar una presencia de géneros 50-50, puesto que esto es algo muy complicado y "no natural" pero, dado que estas diferencias son muy extremas en algunos casos, todavía se perciben diferencias con incidencia de otros factores más allá de la diferenciación natural.

También en torno a estas primeras diferencias vistas en las condiciones laborales de los primeros empleos, mediante una mayor profundización que no ha sido llevada a cabo en este trabajo, sería necesario un mayor conocimiento y, con este, actuaciones para su control.

Sobre la formación continuada, quizá enfocada esta línea actuación a empleadores y empresas, promover la igualdad en cuanto a aspectos formativos en la empresa de hombres y mujeres, independientemente del rango, puesto o área de dedicación de la persona. Que las empresas estas sean actúen en cuanto al progreso igualitario de sus empleados y empleadas, siendo conscientes de su contribución a la carrera profesional de las personas.

Finalizando aquí las llamadas propuestas de actuación, sobre otro aspecto diferenciado, cabe la necesidad de hacer hincapié en el difícil acceso a los datos. Todos generamos constantemente datos que, bien usados, pueden contribuir al bien común en muchos aspectos. Pero a la hora de la verdad, basando esto en la propia experiencia de este trabajo, es difícil acceder en gran parte a estos datos, se encuentran accesos muy restringidos o inaccesibles de manera gratuita. Si bien es cierto que hay muchos datos sensibles por su procedencia, hay otros muchos a los que se podría tener un acceso más a abierto que a día de hoy no se produce.

Por otro lado, claro está que es difícil abordar una cuestión social tan amplia en su totalidad mediante un trabajo como este y que, a medida que este ha ido avanzando han ido surgiendo nuevas vías a explorar o diferentes enfoques que conocer. Es por esto que, ante la imposibilidad de estudiar y analizar todas las variables que podrían resultar diferentes entre mujeres y hombres, a continuación, se establecerán nuevas líneas de investigación, con la intención de que estas puedan ser estudiadas en otro momento y, de esta manera, poder seguir avanzando hacia la igualdad, gracias al conocimiento de la realidad mediante los datos.

Uno de los aspectos que hubiera sido muy interesante de analizar es la conciliación laboral y familiar, un tema que últimamente ha salido más a relucir en los medios de comunicación y debates políticos y, en consecuencia, está más latente en la sociedad. En torno a este sería interesante analizar la repercusión de la cohesión entre vida familiar y laboral en la brecha salarial. De esta manera, seguir ampliando la visión de la vida profesional de una persona.

Más en relación con este trabajo, se establece otra línea de investigación clara en torno a los sectores profesionales. Aunque en el trabajo se ha intentado analizar cómo se comportan las diferencias de género dentro de cada uno, se cree que sería interesante adentrarse más en el estudio detallado de estos, juntando un análisis de género y de sectores, para así llegar a puntos más clave en cada uno de ellos, puesto que generalizar en según qué, teniendo en

cuenta las diferencias latentes que existen actualmente entre diferencias áreas, puede ser útil para una visión general, pero quizá limitado para la toma de decisiones concretas.

Sobre aspectos metodológicos llevados a cabo, comentar aquí los buenos resultados obtenidos mediante el cálculo de los clústers y las reglas de asociación. Aunque quizá en este trabajo no se les ha dado toda la importancia o relevancia posible, a partir de este trabajo sería favorable la continuación de aplicación de metodologías como estas, quizá no tan empleadas en sectores sociológicos, pero quedando probados los buenos resultados obtenidos mediante estos.

En cuanto al ámbito de redes sociales, como ya se ha comentado durante el trabajo, también se establece como pendiente el análisis de LinkedIn. En términos laborales y de actualidad, esta red social podría proporcionar datos muy útiles para el posterior análisis de diferencias de interacción entre hombres y mujeres, viendo como estos se comportan en un entorno laboral social y tratando de observar si en este aspecto ellos y ellas son diferentes, o no. Dejando constancia del explicito deseo de acceso a estos datos, lo cual se seguirá intentando.

Bibliografía

- [1] B. C.-R. J. y. I. M. d. A. Anghel, «Brechas Salariales de Género en España,» FEDEA Estudios de Economía Española, Madrid, 2018.
- [2] J. A. Rodrigo, «RPubs,» Enero 2016. [En línea]. Available: https://rpubs.com/Joaquin_AR/220579.
- [3] M. R. Fredy Mendivelso, «Prueba Chi-Cuadrado de independencia,» *Rev.Medica.Sanitas*, vol. 21, nº 2, pp. 92-95, 2018.
- [4] C. Laguna, «Inferencia paramétrica: relación entre variables cualitativas y cuantitativas,» de *Metodología en Salud Pública*, Instituto Aragonés de Ciencias de la Salud.
- [5] J. A. Gallardo, «Universidad de Granada - Análisis de Datos Multivariantes,» [En línea]. Available: https://www.ugr.es/~gallardo/pdf/prog-amp_am.pdf.
- [6] A. Grané, «Distancias estadísticas y escalado multidimensional,» Universidad Carlos III de Madrid.
- [7] B. S. Villanueva, «Pàgina de les assignatures impartides per Karina Gibert,» [En línea]. Available: <https://www-eio.upc.edu/~karina/datamining/refmaterial/AssociationRules/assocRuBea.pdf>.
- [8] M. Galarnyk, «Medium - Towards Data Science,» 29 Oct 2018. [En línea]. Available: <https://towardsdatascience.com/access-data-from-twitter-api-using-r-and-or-python-b8ac342d3efe>.
- [9] J. A. Rodrigo, «Rpubs,» Diciembre 2017. [En línea]. Available: https://rpubs.com/Joaquin_AR/334526.
- [10] D. Calvo, «Diego Calvo,» 17 nov 2016. [En línea]. Available: <http://www.diegocalvo.es/obtener-datos-de-twitter-con-r-usando-su-api/>.
- [11] A. P. Ezequiel Uriel, Introducción al análisis de series temporales, Madrid: Ediciones Paraninfo, 2000.
- [12] T. D. Gauthier, «Detecting Trends Using Spearman's Rank Correlation Coefficient,» *Enviromental Forensics*, vol. 2, pp. 359-362, 2001.
- [13] A. e. a. Ariño, «Via Universitària: Accés, condicions d'aprenentatge, expectatives i retorns dels estudis universitaris (2017-2019),» Xarxa Vives d'Universitats, 2019.
- [14] M. L. Daniela Del Boca, «The Determinants of Motherhood and Work Status: A survey,» IZA, Germany, 2006.
- [15] I. M. d. A. J. Ignacio Conde-Ruiz, «Gender Gaps in the Spanish Labor Market,» FEDEA, Madrid, 2016.

- [16] L. M.-. K. Francine D. Blau, «Understanding international differences in the gender pay gap,» National Bureau of Economic Research, Cambridge, 2001.
- [17] H. B. Nielsen,
«http://web.econ.ku.dk/metrics/Econometrics2_05_II/Slides/08_unitroottests_2pp.pdf,» 2005.
[En línea]. Available:
http://web.econ.ku.dk/metrics/Econometrics2_05_II/Slides/08_unitroottests_2pp.pdf.
- [18] K. T. Rich, «RPubs,» 2017. [En línea]. Available: <https://rpubs.com/richkt/269797>.
- [19] J. A. Rodrigo, «RPubs,» Septiembre 2017. [En línea]. Available:
https://rpubs.com/Joaquin_AR/310338.
- [20] M. S.-M. V. C. Karina Gibert, «Choosing the Right Data Mining Technique,» de *International Congress on Environmental Modelling and Software*, 2010.
- [21] A. Kassambara, «Sthada,» 17 Nov 2017. [En línea]. Available:
<http://www.sthda.com/english/articles/32-r-graphics-essentials/128-plot-time-series-data-using-ggplot/>.
- [22] B. Edullantes, «RPubs,» 2019. [En línea]. Available:
<http://rpubs.com/bedullantes/ggplottimeseries>.

Anexos

Preprocessing

Formación

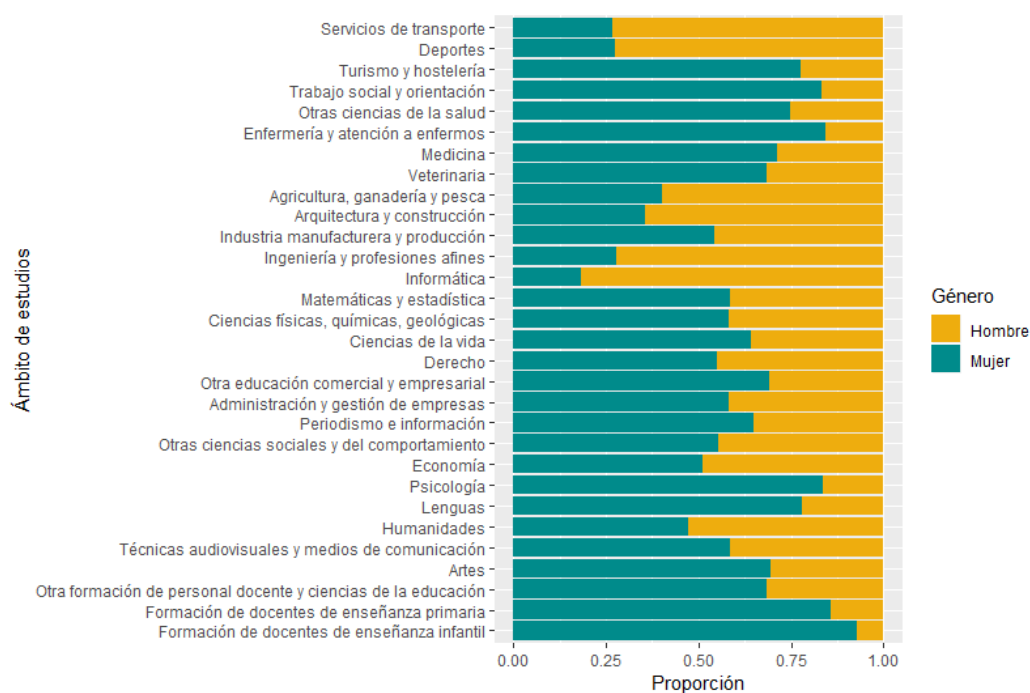
En el capítulo de formación el preprocesamiento de los datos ha consistido en, inicialmente la selección de las variables de estudio entre más de 200. Posteriormente se realizó la asignación a las variables de cada tipo, así como la factorización de todas las variables categóricas. En el caso del nivel de estudios, se realizó una recategorización para unificar las dos variables que recogían este hecho en todos los individuos y en los estudiantes actuales, para la mejor comparación de estas.

Inserción laboral

En el segundo capítulo se siguieron los pasos anteriores sobre preprocesamiento de datos, eliminando las variables no necesarias de la base de datos y categorizando las variables categóricas. En estos datos también se fusionaron dos variables en referencia a la realización de prácticas. Una de ellas recogía la realización de prácticas contempladas en el plan de estudios y otra fuera del plan de estudios, con la intención de analizar la realización de prácticas en general, estas dos se unieron, considerándose sí cuando en una de las dos se había dicho sí.

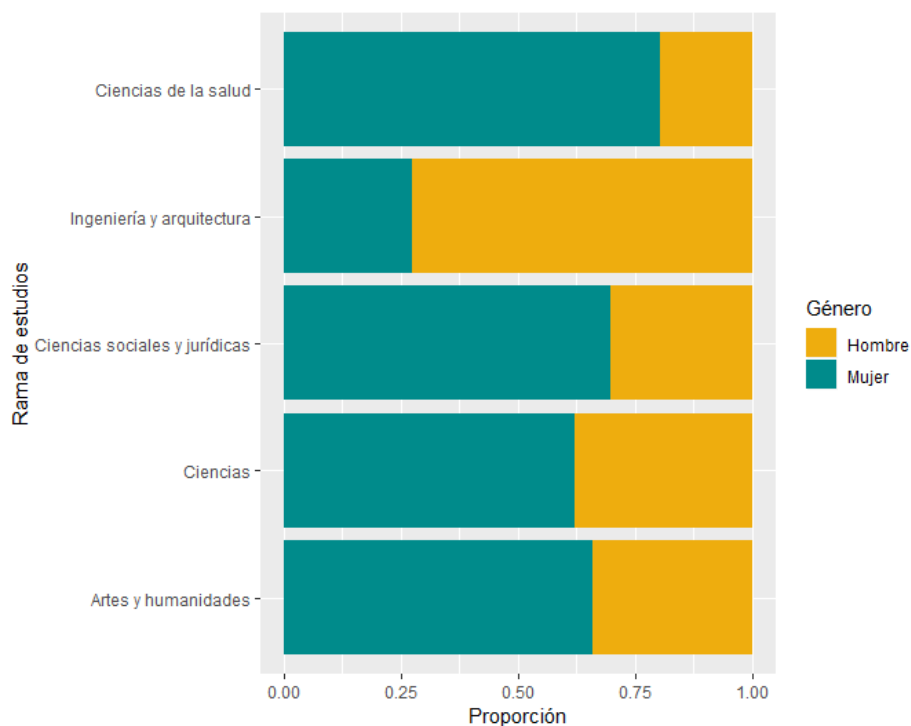
También en estos datos se realizó una selección de categorías para el filtrado y análisis posterior. Inicialmente en los datos había dos variables en referencia al sector, una sobre el ámbito concreto de dedicación y otra sobre la rama en general. Ambas se pueden ver a continuación.

Anexo 1. Ámbito de profesión según género



Fuente: Elaboración propia

Anexo 2. Rama de profesión según género



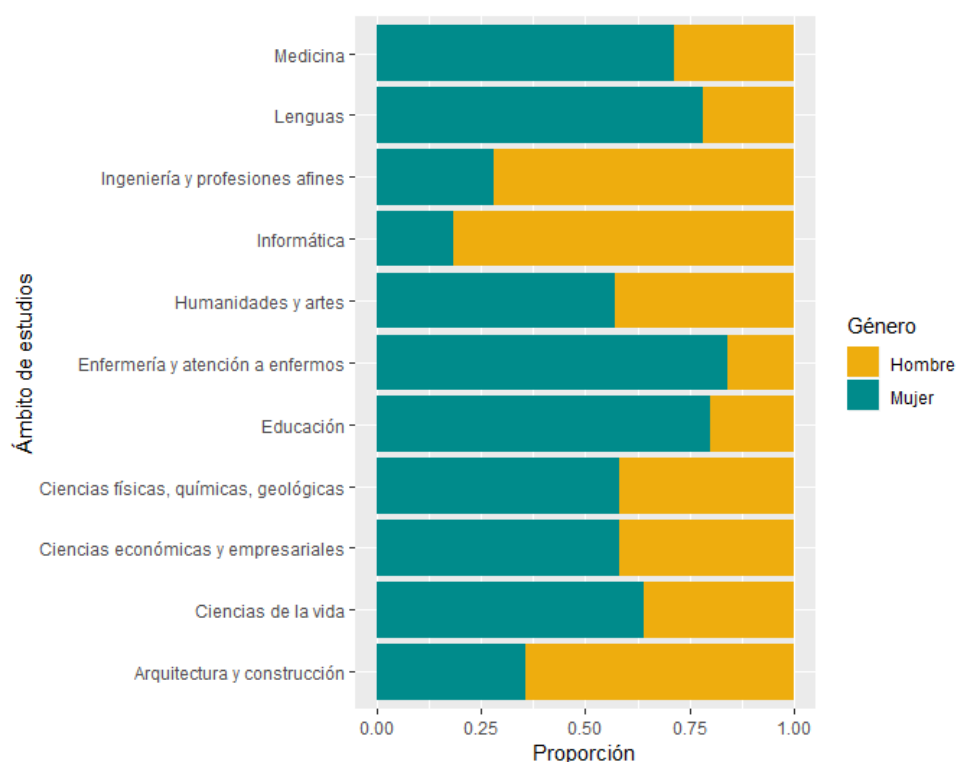
Fuente: Elaboración propia

Estas dos variables se planteaban difíciles de analizar juntamente con otras por dos motivos. En el caso del ámbito, esta variable se encuentra muy desagregada, cosa que podría dificultar el análisis y la interpretación posterior de las técnicas de reglas de asociación o simplemente el análisis cruzado. Por otro lado, en el caso de la rama, esta se encuentra muy agrupada y podría “esconder” fenómenos que se pudieran producir dentro de cada una de las categorías, puesto que agrupan muchas disciplinas con comportamientos diferentes en lo que respecta al género.

Es por esto, que se realizará el análisis en torno a diversos ámbitos concretos, reduciendo así por un lado el gran número de categorías que podrían dificultar el análisis y, por otro lado, pudiendo estudiar al detalle que ocurre en determinadas áreas. Esto se hará escogiendo dos ámbitos por rama, aquellas con mayor número de individuo o bien aquellas con una importancia “histórica” mayor. También en algunos casos se han unido diversos ámbitos en una misma categoría, recogiendo así mayor parte de la muestra inicial.

La variable resultante es la siguiente.

Anexo 3. Variable ámbito de profesión utilizada para el análisis según género
(base: 20452)



Fuente: Elaboración propia

Finalmente, se han cogido las categorías presentadas en el gráfico de la variable ámbito inicial, y se han creado las categorías *educación*, uniendo las categorías del ámbito en la primera variable; y *ciencias económicas y empresariales*, uniendo Economía, Administración y dirección

de empresas y otra educación comercial y empresarial. La base de datos final contiene 20452 individuos, con los que se trabajará para todo el análisis posterior.

Gráficos

Al ser una encuesta, el flujo de las preguntas es muy variable para cada individuo. Con la intención de centrar la atención en los casos válidos de cada variable, cuando haya una elevada aparición de casos con categoría "No aplicable", esta categoría se eliminará de la representación. Cuando no se indique ninguna base, significará que se han utilizado todos los casos de estudio, de otra manera, se indicará en base a cuantos individuos se ha realizado el gráfico.

Resultados R

Anexo 4. Resultado R de la función catdes() para clústering jerárquico var. Generales y sexo

\$'Clúster 1'	Cla/Mod	Mod/Cla	Global	p.value	v.test
DROPEduc=Si	99.724138	20.689655	10.686123	0.000000e+00	Inf
SEX=Hombre	99.888730	89.912720	46.363033	0.000000e+00	Inf
HATLEVEL=Mecánica, electrónica; industria y construcción	89.712389	23.207898	13.324490	5.072526e-303	37.207967
HATLEVEL=Tecnologías de la información y las comunicaciones (TIC)	75.141243	5.708971	3.913332	5.369329e-30	11.378205
HATLEVEL_cat=Formación profesional de grado superior	59.749639	17.756474	15.306950	2.572449e-16	8.191848
HATLEVEL=Agricultura, ganadería, pesca, silvicultura y veterinaria	74.634146	2.189154	1.510797	9.320263e-12	6.816627
HATLEVEL_cat=Segunda etapa de educación secundaria (bachillerato y similares)	53.578826	33.309486	32.021520	9.179598e-04	3.314534
HATLEVEL=Sectores desconocidos o no especificados	55.281690	6.739162	6.279018	2.273399e-02	2.277876
HATLEVEL_cat=Segunda etapa de educación secundaria (formación profesional)	53.794103	15.925025	15.247992	2.374102e-02	2.261295
HATLEVEL=Formación general/básica	53.854167	14.794677	14.149901	2.633644e-02	2.221215
HATLEVEL=Ciencias sociales, periodismo y documentación	47.457627	4.406925	4.782961	3.457730e-02	-2.113274
HATLEVEL=Servicios	47.161572	4.635856	5.063011	1.945065e-02	-2.336779
HATLEVEL_cat=Máster y Doctorado	47.203140	6.882244	7.509765	4.284616e-03	-2.856410
HATLEVEL_cat=Grado universitario (+240 cr. ECTS), licenciatura	47.648579	13.192159	14.260447	2.464702e-04	-3.665900
HATLEVEL=Artes, humanidades y lenguas	43.636364	6.867935	8.106714	5.079895e-08	-5.448491
HATLEVEL_cat=Grado universitario (240 cr. ECTS), diplomatura	42.561205	12.934612	15.653327	2.537969e-19	-8.987119
DROPEduc=No aplicable	47.398190	47.961082	52.118800	1.567611e-23	-9.997204
HATLEVEL=Educación	31.455898	4.235227	6.934925	6.886726e-38	-12.867184
HATLEVEL=Negocios, administración y derecho	39.992135	14.551438	18.741248	3.937414e-38	-12.910304
DROPEduc=No	43.314763	31.148948	37.040312	1.188129e-48	-14.658533
HATLEVEL=Salud y servicios sociales	28.061582	5.737588	10.531358	2.125131e-80	-18.988449
SEX=Mujer	9.686727	10.087280	53.636967	0.000000e+00	-Inf
\$'Clúster 2'	Cla/Mod	Mod/Cla	Global	p.value	v.test
SEX=Mujer	90.3132729	99.89361702	53.636967	0.000000e+00	Inf
HATLEVEL=Salud y servicios sociales	71.9384185	15.62310030	10.531358	2.125131e-80	18.988449
DROPEduc=No	56.6852368	43.29787234	37.040312	1.188129e-48	14.658533
HATLEVEL=Negocios, administración y derecho	60.0078647	23.19148936	18.741248	3.937414e-38	12.910304
HATLEVEL=Educación	68.5441020	9.80243161	6.934925	6.886726e-38	12.867184
DROPEduc=No aplicable	52.6018100	56.53495441	52.118800	1.567611e-23	9.997204
HATLEVEL_cat=Grado universitario (240 cr. ECTS), diplomatura	57.4387947	18.54103343	15.653327	2.537969e-19	8.987119
HATLEVEL=Artes, humanidades y lenguas	56.3636364	9.42249240	8.106714	5.079895e-08	5.448491
HATLEVEL_cat=Grado universitario (+240 cr. ECTS), licenciatura	52.3514212	15.39513678	14.260447	2.464702e-04	3.665900
HATLEVEL_cat=Máster y Doctorado	52.7968597	8.17629179	7.509765	4.284616e-03	2.856410
HATLEVEL=Servicios	52.8384279	5.51671733	5.063011	1.945065e-02	2.336779
HATLEVEL=Ciencias sociales, periodismo y documentación	52.5423729	5.18237082	4.782961	3.457730e-02	2.113274
HATLEVEL=Formación general/básica	46.1458333	13.46504559	14.149901	2.633644e-02	-2.221215
HATLEVEL_cat=Segunda etapa de educación secundaria (formación profesional)	46.2058966	14.52887538	15.247992	2.374102e-02	-2.261295
HATLEVEL=Sectores desconocidos o no especificados	44.7183099	5.79027356	6.279018	2.273399e-02	-2.277876
HATLEVEL_cat=Segunda etapa de educación secundaria (bachillerato y similares)	46.4211738	30.65349544	32.021520	9.179598e-04	-3.314534
HATLEVEL=Agricultura, ganadería, pesca, silvicultura y veterinaria	25.3658537	0.79027356	1.510797	9.320263e-12	-6.816627
HATLEVEL_cat=Formación profesional de grado superior	40.2503611	12.70516717	15.306950	2.572449e-16	-8.191848
HATLEVEL=Tecnologías de la información y las comunicaciones (TIC)	24.8587571	2.00607903	3.913332	5.369329e-30	-11.378205
HATLEVEL=Mecánica, electrónica; industria y construcción	10.2876106	2.82674772	13.324490	5.072526e-303	-37.207967
DROPEduc=Si	0.2758621	0.06079027	10.686123	0.000000e+00	-Inf
SEX=Hombre	0.1112701	0.10638298	46.363033	0.000000e+00	-Inf

Anexo 5. Resultado R de la función catdes() para clúster jerárquico var. Generales sin sexo

\$ 1`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
DROPEduc=No aplicable	83.63970588	97.73628553	52.1188002	0.000000e+00	Inf
HATLEVEL_cat=Segunda etapa de educación secundaria (bachillerato y similares)	60.36823936	43.34104428	32.0215196	6.632654e-142	25.361959
HATFIELD= Sectores desconocidos o no especificados	74.06103286	10.42630535	6.2790184	1.705069e-72	18.007381
HATLEVEL_cat=Segunda etapa de educación secundaria (formación profesional)	60.80231996	20.78651685	15.2479917	4.479974e-58	16.065120
HATFIELD=Formación general/básica	61.40625000	19.48116325	14.1499005	2.782824e-57	15.951467
HATFIELD=Agricultura, ganadería, pesca, silvicultura y veterinaria	84.87804878	2.87508262	1.5107967	2.648006e-33	12.024307
HATLEVEL_cat=Formación profesional de grado superior	53.58690419	18.39061467	15.3069497	4.562555e-19	8.922409
HATLEVEL_cat=Grado universitario (+240 cr. ECTS), licenciatura	52.09302326	16.65565102	14.2604466	9.532167e-13	7.137098
SEX=Hombre	47.84612939	49.73562459	46.3630334	1.574009e-12	7.067807
DROPEduc=Negativa/ No sabe	95.23809524	0.33046927	0.1547645	1.196179e-06	4.856269
HATFIELD=Mecánica, electrónica; industria y construcción	49.28097345	14.72240582	13.3244896	1.794947e-05	4.288981
HATFIELD=Ciencias naturales, químicas, físicas y matemáticas	49.88938053	7.45208196	6.6622448	9.650068e-04	3.300535
HATFIELD=Negocios, administración y derecho	42.35155328	17.79576999	18.7412484	1.125413e-02	-2.534707
HATFIELD=Artes, humanidades y lenguas	40.54545455	7.36946464	8.1067138	4.657400e-03	-2.829821
HATFIELD=Servicios	38.42794760	4.36219432	5.0630113	7.926523e-04	-3.355348
HATFIELD=Ciencias sociales, periodismo y documentación	32.35747304	3.46992730	4.7829612	7.393687e-11	-6.512446
HATFIELD=Tecnologías de la información y las comunicaciones (TIC)	30.50847458	2.67680106	3.9133319	1.287873e-11	-6.769995
SEX=Mujer	41.79719703	50.26437541	53.6369666	1.574009e-12	-7.067807
HATFIELD=Salud y servicios sociales	27.50174948	6.49372108	10.5313582	9.482618e-45	-14.035276
HATFIELD=Educación	18.49096706	2.87508262	6.9349252	3.035886e-68	-17.457175
HATLEVEL_cat=Máster y Doctorado	0.09813543	0.01652346	7.5097649	1.637772e-273	-35.336126
DROPEduc=No	2.30799841	1.91672174	37.0403125	0.000000e+00	-Inf
DROPEduc=Si	0.06896552	0.01652346	10.6861228	0.000000e+00	-Inf
HATLEVEL_cat=Grado universitario (240 cr. ECTS), diplomatura	2.30696798	0.80964970	15.6533274	0.000000e+00	-Inf

\$ 2`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
DROPEduc=No	97.692002	80.9296193	37.0403125	0.000000e+00	Inf
HATLEVEL_cat=Grado universitario (240 cr. ECTS), diplomatura	86.158192	30.1631778	15.6533274	0.000000e+00	Inf
HATLEVEL_cat=Máster y Doctorado	87.340530	14.6695237	7.5097649	1.533971e-191	29.521309
HATFIELD=Educación	70.031881	10.8620405	6.9349252	2.409664e-59	16.245349
HATFIELD=Salud y servicios sociales	61.651505	14.5211802	10.5313582	4.963847e-42	13.584281
SEX=Mujer	48.502336	58.1836163	53.6369666	1.186694e-21	9.559201
HATFIELD=Ciencias sociales, periodismo y documentación	55.624037	5.9502225	4.7829612	1.160220e-08	5.705471
HATFIELD=Negocios, administración y derecho	47.856862	20.0593374	18.7412484	4.114235e-04	3.532646
HATFIELD=Tecnologías de la información y las comunicaciones (TIC)	50.659134	4.4338223	3.9133319	5.080219e-03	2.801903
HATFIELD=Ciencias naturales, químicas, físicas y matemáticas	41.150442	6.1315312	6.6622448	2.554675e-02	-2.233033
HATFIELD=Mecánica, electrónica; industria y construcción	41.039823	12.2300972	13.3244896	7.236996e-04	-3.804400
DROPEduc=Negativa/ No sabe	0.000000	0.0000000	0.1547645	3.888393e-06	-4.617260
HATLEVEL_cat=Grado universitario (+240 cr. ECTS), licenciatura	39.483204	12.5927147	14.2604466	5.340804e-07	-5.013647
SEX=Hombre	40.327452	41.8163837	46.3630334	1.186694e-21	-9.559201
HATFIELD=Agricultura, ganadería, pesca, silvicultura y veterinaria	5.853659	0.1977913	1.5107967	5.026688e-36	-12.531401
HATLEVEL_cat=Formación profesional de grado superior	31.439576	10.7631449	15.3069497	6.637728e-41	-13.393072
HATLEVEL_cat=Segunda etapa de educación secundaria (formación profesional)	30.352827	10.3510796	15.2479917	1.600685e-47	-14.480866
HATFIELD= Sectores desconocidos o no especificados	21.244131	2.9833526	6.2790184	2.856685e-49	-14.755005
HATFIELD=Formación general/básica	28.645833	9.0654360	14.1499005	1.468380e-54	-15.555138
HATLEVEL_cat=Segunda etapa de educación secundaria (bachillerato y similares)	29.965478	21.4603593	32.0215196	3.633384e-127	-23.988874
DROPEduc=No aplicable	16.360294	19.0703807	52.1188002	0.000000e+00	-Inf
DROPEduc=Si	0.000000	0.0000000	10.6861228	0.000000e+00	-Inf

\$ 3`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
DROPEduc=Si	99.931034	99.931034	10.686123	0.000000e+00	Inf
HATLEVEL_cat=Formación profesional de grado superior	14.973519	21.448276	15.306950	4.265597e-11	6.594546
HATFIELD=Tecnologías de la información y las comunicaciones (TIC)	18.832392	6.896552	3.913332	1.221262e-08	5.696731
SEX=Hombre	11.826419	51.310345	46.363033	6.589098e-05	3.990655
HATFIELD=Servicios	15.429403	7.310345	5.063011	8.747664e-05	3.922935
HATFIELD=Artes, humanidades y lenguas	13.454545	10.206897	8.106714	2.604768e-03	3.010897
HATLEVEL_cat=Máster y Doctorado	12.561335	8.827586	7.509765	4.779170e-02	1.979216
HATLEVEL_cat=Segunda etapa de educación secundaria (bachillerato y similares)	9.666283	28.965517	32.021520	7.909686e-03	-2.655901
HATLEVEL_cat=Segunda etapa de educación secundaria (formación profesional)	8.844853	12.620690	15.247992	2.687193e-03	-3.001425
HATLEVEL_cat=Grado universitario (+240 cr. ECTS), licenciatura	8.423773	11.241379	14.260447	3.597936e-04	-3.567944
SEX=Mujer	9.700467	48.689655	53.636967	6.589098e-05	-3.990655
HATFIELD= Sectores desconocidos o no especificados	4.694836	2.758621	6.279018	1.265651e-10	-6.431242
DROPEduc=No	0.000000	0.000000	37.040312	6.590893e-314	-37.875196
DROPEduc=No aplicable	0.000000	0.000000	52.118800	0.000000e+00	-Inf

Anexo 6. Resultado R de la función catdes() para clúster jerárquico estudiantes actuales

```
$`1`
SEX=Mujer
FEDFIELD=Salud y servicios sociales
FEDLEVEL_cat=Grados universitarios(+240 ECTS) / Licenciados
FEDPAID=Pagado totalmente por usted
FEDFIELD=Educación
FEDFIELD=Artes, humanidades y lenguas
FEDLEVEL_cat=Máster, Ciencias de la Salud y Doctorado
FEDLEVEL_cat=Grado universitario (240 cr. ECTS), diplomatura
FEDFIELD=Ciencias sociales, periodismo y documentación
FEDWORKTIME=Sólo fuera de horas de trabajo remuneradas
FEDFIELD=Negocios, administración y derecho
FEDFIELD=Servicios
FEDWORKTIME=Mayoritariamente durante horas de trabajo remuneradas
FEDFIELD=Agricultura, ganadería, pesca, silvicultura y veterinaria
FEDPAID=Pagado totalmente por otra persona o institución
FEDWORKTIME=Sólo durante horas de trabajo remuneradas
FEDLEVEL_cat=Educación primaria o inferior
FEDFIELD=Formación general/básica
FEDPAID=La actividad era gratuita
FEDFIELD=Tecnologías de la información y las comunicaciones (TIC)
FEDFIELD=Mecánica, electrónica ; industria y construcción
FEDFIELD=No aplicable
FEDLEVEL_cat=Primera etapa de educación secundaria
SEX=Hombre
Cla/Mod Mod/Cla Global p.value v.test
91.875682 95.0902935 53.3139535 0.000000e+00 Inf
83.295711 20.8239278 12.8779070 2.357475e-50 14.922370
87.323944 10.4966140 6.1918605 3.772622e-30 11.408948
57.291667 52.7652370 47.4418605 1.132878e-10 6.448064
71.365639 9.1422122 6.5988372 3.364111e-10 6.280996
59.649123 24.9435666 21.5406977 5.332580e-07 5.013943
59.840426 12.6975169 10.9302326 6.033494e-04 3.430104
55.157438 28.6681716 26.7732558 9.681970e-03 2.586985
60.209424 6.4898420 5.5523256 1.327039e-02 2.476429
54.769231 30.1354402 28.3430233 1.621824e-02 2.403967
55.804481 15.4627540 14.2732558 3.985891e-02 2.055208
44.444444 5.8690745 6.8023256 2.535367e-02 -2.235971
37.500000 1.3544018 1.8604651 2.427759e-02 -2.252708
32.692308 0.9593679 1.5116279 6.365221e-03 -2.728349
46.254072 24.0406321 26.7732558 1.928447e-04 -3.728210
22.448980 0.6207675 1.4244186 3.342300e-05 -4.148795
0.000000 0.0000000 0.6686047 5.438814e-08 -5.436333
30.487805 2.8216704 4.7674419 2.718367e-08 -5.558671
35.777126 6.8848758 9.9127907 8.066316e-10 -6.143617
21.142857 2.0880361 5.0872093 3.128991e-17 -8.441646
25.316456 4.5146727 9.1860465 3.072410e-23 -9.930329
0.000000 0.0000000 5.3488372 5.859895e-61 -16.471701
2.489627 0.3386005 7.0058140 1.089364e-68 -17.515597
5.417186 4.9097065 46.6860465 0.000000e+00 -Inf

$`2`
SEX=Hombre
FEDLEVEL_cat=Primera etapa de educación secundaria
FEDFIELD=No aplicable
FEDFIELD=Mecánica, electrónica ; industria y construcción
FEDFIELD=Tecnologías de la información y las comunicaciones (TIC)
FEDPAID=La actividad era gratuita
FEDFIELD=Formación general/básica
FEDLEVEL_cat=Educación primaria o inferior
FEDWORKTIME=Sólo durante horas de trabajo remuneradas
FEDPAID=Pagado totalmente por otra persona o institución
FEDFIELD=Agricultura, ganadería, pesca, silvicultura y veterinaria
FEDWORKTIME=Mayoritariamente durante horas de trabajo remuneradas
FEDFIELD=Servicios
FEDFIELD=Negocios, administración y derecho
FEDWORKTIME=Sólo fuera de horas de trabajo remuneradas
FEDFIELD=Ciencias sociales, periodismo y documentación
FEDLEVEL_cat=Grado universitario (240 cr. ECTS), diplomatura
FEDLEVEL_cat=Máster, Ciencias de la Salud y Doctorado
FEDFIELD=Artes, humanidades y lenguas
FEDFIELD=Educación
FEDPAID=Pagado totalmente por usted
FEDLEVEL_cat=Grados universitarios(+240 ECTS) / Licenciados
FEDFIELD=Salud y servicios sociales
SEX=Mujer
Cla/Mod Mod/Cla Global p.value v.test
94.582814 91.067146 46.6860465 0.000000e+00 Inf
97.510373 14.088729 7.0058140 1.089364e-68 17.515597
100.000000 11.031175 5.3488372 5.859895e-61 16.471701
74.683544 14.148681 9.1860465 3.072410e-23 9.930329
78.857143 8.273381 5.0872093 3.128991e-17 8.441646
64.222874 13.129496 9.9127907 8.066316e-10 6.143617
69.512195 6.834532 4.7674419 2.718367e-08 5.558671
100.000000 1.378897 0.6686047 5.438814e-08 5.436333
77.551020 2.278177 1.4244186 3.342300e-05 4.148795
53.745928 29.676259 26.7732558 1.928447e-04 3.728210
67.307692 2.098321 1.5116279 6.365221e-03 2.728349
62.500000 2.398082 1.8604651 2.427759e-02 2.252708
55.555556 7.793765 6.8023256 2.535367e-02 2.235971
44.195519 13.009592 14.2732558 3.985891e-02 -2.055208
45.230769 26.438849 28.3430233 1.621824e-02 -2.403967
39.790576 4.556355 5.5523256 1.327039e-02 -2.476429
44.842562 24.760192 26.7732558 9.681970e-03 -2.586985
40.159574 9.052758 10.9302326 6.033494e-04 -3.430104
40.350877 17.925659 21.5406977 5.332580e-07 -5.013943
28.634361 3.896882 6.5988372 3.364111e-10 -6.280996
42.708333 41.786571 47.4418605 1.132878e-10 -6.448064
12.676056 1.618705 6.1918605 3.772622e-30 -11.408948
16.704289 4.436451 12.8779070 2.357475e-50 -14.922370
8.124318 8.932854 53.3139535 0.000000e+00 -Inf
```

Scripts

```
#####
# DEPURACIO DADES EDUCACIO #
#####

dd <- df[, c(1:12,19:22,24:36,38:141,182:198,215:236,240:247)]

#t(colnames(df)) et dona el nom de la variable i la seva posicio

v <- c()

for(i in 1:ncol(dd)){

  if(class(dd[,i])=="character"){

    v <- c(v, i)

  }

}
```

```

length(v)
v

#for(i in v){
#  dd[,i] <- as.factor(dd[,i])
#}

levels(dd[,2])=c("Andaluc?a", "Arag?n", "Asturias", "Illes
Balears", "Canarias", "Cantabria", "Castilla y Le?n", "Castilla-La
Mancha", "Catalu?a", "Comunitat
Valenciana", "Extremadura", "Galicia", "Madrid", "Murcia", "Navarra", "Pa?s
Vasco", "La Rioja", "Ceuta", "Melilla")

levels(dd[,3])=c("Febrero", "Marzo")

levels(dd[,5])=c("Hombre", "Mujer")

levels(dd[,6])=c("Enero", "Febrero", "Marzo", "Abril", "Mayo", "Junio", "Julio", "
Agosto", "Septiembre", "Octubre", "Noviembre", "Diciembre")

levels(dd[,9])=c("Austria", "B?lgica", "Bulgaria", "Espa?a", "Francia", "Hungr?a
", "Irlanda", "Italia", "Noruega", "Pa?ses
Bajos", "Polonia", "Portugal", "Andorra", "Reino
Unido", "Alemania", "Ruman?a", "Suecia", "Suiza",
"Ucrania", "Letonia", "Moldavia", "Belar?s", "Georgia", "Estonia", "Lituania", "Ar
menia", "Rusia", "Otros pa?ses de
Europa", "Argelia", "Camer?n", "Egipto", "Gambia", "Ghana", "Guinea", "Guinea-
Bissau", "Guinea
Ecuatorial", "Libia", "Mali", "Marruecos", "Mauritania", "Mozambique", "Nigeria",
"Senegal", "Congo", "Estados Unidos", "M?xico", "Belice", "Costa
Rica", "Cuba", "El Salvador", "Guatemala", "Honduras", "Nicaragua", "Rep?blica
Dominicana", "Argentina", "Bolivia", "Brasil", "Colombia", "Chile", "Ecuador", "Pa
raguay", "Per?", "Uruguay", "Venezuela", "Bangladesh", "China", "Filipinas", "Indi
a", "Israel", "Jap?n", "Jordania", "Mongolia", "Pakist?n", "Corea", "Siria", "Taila
ndia", "Uzbekist?n", "Ap?tridas", "Negativa/No sabe")

levels(dd[,10])=c("S?", "No")

levels(dd[,11])=c("Austria", "B?lgica", "Bulgaria", "Francia", "Grecia", "Hungr?
a", "Irlanda", "Italia", "Luxemburgo", "Noruega", "Pa?ses
Bajos", "Polonia", "Portugal", "Andorra", "Reino
Unido", "Alemania", "Ruman?a", "Suecia", "Suiza",
"Ucrania", "Letonia", "Moldavia", "Belar?s", "Georgia", "Lituania", "Croacia", "Ar
menia", "Rusia", "Angola", "Argelia", "Cabo
Verde", "Camer?n", "Egipto", "Gambia", "Ghana", "Guinea", "Guinea-Bissau", "Guinea
Ecuatorial", "Libia", "Mali", "Marruecos", "Mauritania", "Mozambique", "Nigeria",
"Senegal", "T?nez", "Congo", "Otros pa?ses de ?frica", "Canad?", "Estados
Unidos", "M?xico", "Costa Rica", "Cuba", "El
Salvador", "Guatemala", "Honduras", "Nicaragua", "Rep?blica
Dominicana", "Argentina", "Bolivia", "Brasil", "Colombia", "Chile", "Ecuador", "Pa
raguay", "Per?", "Uruguay", "Venezuela", "Otros pa?ses de Am?rica",
"Bangladesh", "China", "Filipinas", "India", "Ir?n", "Israel", "Jap?n", "Jordania",
"Laos", "L?bano", "Mongolia", "Pakist?n", "Corea", "Siria", "Tailandia", "Uzbekis
t?n", "Australia", "Negativa/No sabe", "No aplicable")

levels(dd[,12])=c("1 a?o o menos", "M?s de 1/2 a?os", "M?s de 2/3 a?os",
"M?s de 3/4 a?os", "M?s de 4/5 a?os", "M?s de 5/6 a?os", "M?s de 6/7

```

```
a?os", "M?s de 7/8 a?os", "M?s de 8/9 a?os", "M?s de 9/10 a?os", "M?s de
10 a?os", "Negativa/ No sabe", "No aplicable")
```

```
levels(dd[,13])=c("Jornada parcial", "Jornada completa", "Desempleado",
"Estudiante", "Jubilado", "Incapacitado permanente", "Dedicado a las
labores del hogar", "Otra situacion", "Negativa/ No sabe")
```

```
#JOBISCO
```

```
levels(dd[,14])=c("Directores ejecutivos, personal directivo de la
administraci?n p?blica y miembros del poder ejecutivo y de los cuerpos
legislativos", "Directores administradores y comerciales", "Directores y
gerentes de producci?n y operaciones", "Gerentes de hoteles,
restaurantes, comercios y otros servicios", "Profesionales de las
ciencias y de la ingenier?a", "Profesionales de la salud",
"Profesionales de la ense?anza", "Especialistas en organizaci?n de la
administraci?n p?blica y de empresas", "Profesionales de tecnolog?a de la
informaci?n y las comunicaciones", "Profesionales en derecho, en ciencias
sociales y culturales", "Profesionales de las ciencias y la ingenier?a de
nivel medio", "Profesionales de nivel medio de la salud", "Profesionales de
nivel medio en operaciones financieras y administrativas",
"Profesionales de nivel medio de servicios jur?dicos, sociales, culturales
y afines", "T?cnicos de la tecnolog?a de la informaci?n y las
comunicaciones", "Oficinistas", "Empleados en trato directo con el
p?blico", "Empleados contables y encargados del registro de materiales",
"Otro personal de apoyo administrativo", "Trabajadores de los servicios
personales", "Vendedores", "Trabajadores de los cuidados personales",
"Personal de los servicios de protecci?n", "Agricultores y trabajadores
calificados de explotaciones agropecuarias con destino al mercado",
"Trabajadores forestales calificados, pescadores y cazadores",
"Trabajadores agropecuarios, pescadores, cazadores y recolectores de
subsistencia", "Oficiales y operarios de la construcci?n excluyendo
electricistas", "Oficiales y operarios de la metalurgia, la construcci?n
mec?nica y afines", "Artesanos y operarios de las artes gr?ficas",
"Trabajadores especializados en electricidad y la electrotecnolog?a",
"Operarios y oficiales de procesamiento de alimentos, de la confecci?n,
ebanistas, otros artesanos y afines", "Operadores de instalaciones fijas
y m?quinas", "Ensambladores", "Conductores de veh?culos y operadores de
equipos pesados m?viles", "Limpiadores y asistentes", "Peones
agropecuarios, pesqueros y forestales", "Peones de la miner?a, la
construcci?n, la industria manufacturera y el transporte", "Ayudantes de
preparaci?n de alimentos", "Vendedores ambulantes de servicios y
afines", "Recolectores de desechos y otras ocupaciones elementales",
"Oficiales de las fuerzas armadas", "Suboficiales de las fuerzas armadas",
"Otros miembros de las fuerzas armadas", "Negativa /No sabe", "No
aplicable")
```

```
#LOCNACE -- NO HO FA BE
```

```
#levels(dd[,15])=c("Agricultura, ganader?a, caza y servicios relacionados
con las mismas", "Silvicultura y explotaci?n forestal", "Pesca y
acuicultura", "Extracci?n de antracita, hulla y lignito", "Extracci?n de
crudo de petr?leo y gas natural", "Extracci?n de minerales met?licos",
"Otras industrias extractivas", "Actividades de apoyo a las industrias
extractivas", "Industria de la alimentaci?n", "Fabricaci?n de bebidas",
"Industria del tabaco", "Industria textil", "Confecci?n de prendas de
vestir", "Industria del cuero y del calzado", "Industria de la madera y
del corcho, excepto muebles; cester?a y esparter?a", "Industria del
papel", "Artes gr?ficas y reproducci?n de soportes grabados", "Coquer?as
y refino de petr?leo", "Industria qu?mica", "Fabricaci?n de productos
```

farmac?uticos", "Fabricaci?n de productos de caucho y pl?sticos",
 "Fabricaci?n de otros productos minerales no met?licos", "Metalurgia;
 fabricaci?n de productos de hierro, acero y ferroaleaciones", "Fabricaci?n
 de productos met?licos, excepto maquinaria y equipo", "Fabricaci?n de
 productos inform?ticos, electr?nicos y ?pticos", "Fabricaci?n de
 material y equipo el?ctrico", "Fabricaci?n de maquinaria y equipo
 n.c.o.p.", "Fabricaci?n de veh?culos de motor, remolques y
 semirremolques", "Fabricaci?n de otro material de transporte",
 "Fabricaci?n de muebles", "Otras industrias manufactureras", "Reparaci?n
 e instalaci?n de maquinaria y equipo", "Suministro de energ?a el?ctrica,
 gas, vapor y aire acondicionado", "Captaci?n, depuraci?n y distribuci?n de
 agua", "Recogida y tratamiento de aguas residuales", "Recogida,
 tratamiento y eliminaci?n de residuos; valorizaci?n", "Actividades de
 descontaminaci?n y otros servicios de gesti?n de residuos", "Construcci?n
 de edificios", "Ingenier?a civil", "Actividades de construcci?n
 especializada", "Venta y reparaci?n de veh?culos de motor y
 motocicletas", "Comercio al por mayor e intermediarios del comercio,
 excepto de veh?culos de motor y motocicletas", "Comercio al por menor,
 excepto de veh?culos de motor y motocicletas", "Transporte terrestre y por
 tuber?a", "Transporte mar?timo y por v?as navegables interiores",
 "Transporte a?reo", "Almacenamiento y actividades anexas al transporte",
 "Actividades postales y de correos", "Servicios de alojamiento",
 "Servicios de comidas y bebidas", "Edici?n", "Actividades
 cinematogr?ficas, de v?deo y de programas de televisi?n, grabaci?n de
 sonido y edici?n musical", "Actividades de programaci?n y emisi?n de
 radio y televisi?n", "Telecomunicaciones", "Programaci?n, consultor?a y
 otras actividades relacionadas con la inform?tica", "Servicios de
 informaci?n", "Servicios financieros, excepto seguros y fondos de
 pensiones", "Seguros, reaseguros y fondos de pensiones, excepto Seguridad
 Social obligatoria", "Actividades auxiliares a los servicios financieros y
 a los seguros", "Actividades inmobiliarias", "Actividades jur?dicas y
 de contabilidad", "Actividades de las sedes centrales; actividades de
 consultor?a de gesti?n empresarial", "Servicios t?cnicos de arquitectura
 e ingenier?a; ensayos y an?lisis t?cnicos", "Investigaci?n y desarrollo",
 "Publicidad y estudios de mercado", "Otras actividades profesionales,
 cient?ficas y t?cnicas", "Actividades veterinarias", "Actividades de
 alquiler", "Actividades relacionadas con el empleo", "Actividades de
 agencias de viajes, operadores tur?sticos, servicios de reservas y
 actividades relacionadas con los mismos", "Actividades de seguridad e
 investigaci?n", "Servicios a edificios y actividades de jardiner?a",
 "Actividades administrativas de oficina y otras actividades auxiliares a
 las empresas", "Administraci?n P?blica y defensa; Seguridad Social
 obligatoria", "Educaci?n", "Actividades sanitarias", "Asistencia en
 establecimientos residenciales", "Actividades de servicios sociales sin
 alojamiento", "Actividades de creaci?n, art?sticas y espect?culos",
 "Actividades de bibliotecas, archivos, museos y otras actividades
 culturales", "Actividades de juegos de azar y apuestas", "Actividades
 deportivas, recreativas y de entretenimiento", "Actividades asociativas",
 "Reparaci?n de ordenadores, efectos personales y art?culos de uso
 dom?stico", "Otros servicios personales", "Actividades de los hogares
 como empleadores de personal dom?stico", "Actividades de organizaciones
 y organismos extraterritoriales", "Negativa/ No sabe", "No
 aplicable")

#JOBSTAT

levels(dd[,16]) = c("Empresario con asalariados", "Trabajador aut?nomo o
 empresario sin asalariados", "Asalariado con un trabajo fijo o un contrato
 de duraci?n indefinida", "Asalariado con trabajo temporal o con contrato


```

de trabajo de duraci?n determinada",      "Ayuda en la empresa o negocio
familiar",      "Negativa/ No sabe",      "No aplicable")

#EMP12M
levels(dd[,18]) = c("S?",      "No",      "Negativa/ No sabe",      "No aplicable")

#HATLEVEL
levels(dd[,19]) = c("Analfabetos ", "Estudios primarios incompletos ",
"Educaci?n primaria ", "Primera etapa de educaci?n secundaria sin t?tulo",
"Primera etapa de educaci?n secundaria con t?tulo",      "Certificados de
profesionalidad de nivel 1 o nivel 2", "Bachillerato y similares",
"Ense?anzas de formaci?n profesional",      "Formaci?n Profesional B?sica
(FPB)", "Certificados de profesionalidad de nivel 3",      "Ense?anzas de
formaci?n profesional / T?tulos propios universitarios", "Grados
universitarios(240 ECTS) / Diplomados universitarios", "Grados
universitarios(+240 ECTS) / Licenciados",      "M?steres oficiales
universitarios",      "Doctorado universitario")

#HATFIELD
levels(dd[,20])=c("Formaci?n general/b?sica", "Educaci?n",      "Artes,
humanidades y lenguas", "Ciencias sociales, periodismo y documentaci?n",
"Negocios, administraci?n y derecho",      "Ciencias naturales, qu?micas,
f?sicas y matem?ticas",      "Tecnolog?as de la informaci?n y las
comunicaciones (TIC)",      "Mec?nica, electr?nica; industria y construcci?n",
"Agricultura, ganadera, pesca, silvicultura y veterinaria", "Salud y
servicios sociales",      "Servicios",      "Sectores desconocidos o no
especificados", "Negativa/ No sabe",      "No aplicable")

#DROPEduc
levels(dd[,22])=c("S?", "No",      "Negativa/ No sabe",      "No aplicable")

#DROPEducLEVEL
levels(dd[,23])=c("ESO /EGB / Bachiller Elemental ",      "Certificados de
profesionalidad 1 y 2",      "Profesional de m?sica o danza o EOI",
"Bachillerato / BUP",      "Grado medio", "Certificado de profesionalidad de
nivel 3",      "Grado superior",      "Grado universitario (240 ECTS),
diplomados ", "Graduados universitario (m?s de 240 ECTS), licenciados,
ingenieros y MIR",      "Doctorado ",      "Negativa/No sabe", "No aplicable")

#HATFATHER
levels(dd[,24])=c("Estudios primarios o inferiores (incluido si no sabe
leer o escribir)", "Educaci?n secundaria obligatoria, Graduado Escolar ,
Bachiller elemental o equivalentes.",      "Bachiller, Bachiller Superior,
BUP o Formaci?n Profesional de Grado Medio o equivalente ", "Estudios
Universitarios o Formaci?n Profesional de Grado Superior o equivalente ",
"Negativa /No sabe")

#HATMOTHER
levels(dd[,25])=c("Estudios primarios o inferiores (incluido si no sabe
leer o escribir)", "Educaci?n secundaria obligatoria, Graduado Escolar ,
Bachiller elemental o equivalentes.",      "Bachiller, Bachiller Superior,
BUP o Formaci?n Profesional de Grado Medio o equivalente ", "Estudios
Universitarios o Formaci?n Profesional de Grado Superior o equivalente ",
"Negativa /No sabe")

#FEDLEVEL
levels(dd[,27])=c("Educaci?n b?sica para adultos",      "M?sica y danza",
"E.S.O. (1?-3?)",      "Secundaria de adultos",      "Certificados de

```


profesionalidad de nivel 1", "E.S.O (4º)", "Bachillerato", "Formación profesional, artes plásticas, diseño y deportes grado medio", "Enseñanzas profesionales de música y danza", "EOI", "Cualificación profesional inicial", "Certificados de profesionalidad de nivel 2", "Formación profesional básica", "Certificados de profesionalidad de nivel 3", "Formación profesional, artes plásticas, diseño y deportes grado superior", "Títulos propios universitario", "Grados universitarios (240 ECTS)", "Diplomaturas universitarias", "Títulos propios universitarios de experto o especialista", "Grados universitarios (+ 240 ECTS)", "Licenciaturas y equivalentes", "Másteres oficiales universitarios", "Especialidades en Ciencias de la Salud", "Títulos propios universitarios de máster", "Doctorado ", "No aplicable")

#FEDFIELD

levels(dd[,28])=c("Formación general/básica", "Educación", "Artes, humanidades y lenguas", "Ciencias sociales, periodismo y documentación", "Negocios, administración y derecho", "Ciencias naturales, químicas, físicas y matemáticas", "Tecnologías de la información y las comunicaciones (TIC)", "Mecánica, electrónica ; industria y construcción", "Agricultura, ganadería, pesca, silvicultura y veterinaria", "Salud y servicios sociales", "Servicios", "Sectores desconocidos o no especificados", "Negativa/ No sabe", "No aplicable")

#FEDCOMP

levels(dd[,30])=c("Abandon? los estudios antes de su finalizaci?n", "Todav?a est? cursando estos estudios", "Los ha finalizado", "Negativa/No sabe", "No aplicable")

#FEDIST

levels(dd[,31])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#FEDDISTOL

levels(dd[,32])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#FEDOERA

levels(dd[,33])=c("Con mucha frecuencia", "Frecuentemente", "A veces", "Nunca", "Negativa/No sabe", "No aplicable")

#FEDOERB

levels(dd[,34])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#FEDREASON_01a

levels(dd[,35])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,36])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,37])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,38])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,39])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,40])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,41])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,42])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

```

levels(dd[,43])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,44])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#FEDWORKTIME
levels(dd[,45])=c("S?lo durante horas de trabajo remuneradas ",
"Mayoritariamente durante horas de trabajo remuneradas ",
"Mayoritariamente fuera de horas de trabajo remuneradas ", "S?lo fuera de
horas de trabajo remuneradas ", "No trabajaba en ese momento ", "Negativa
/No sabe", "No aplicable")

#FEDPAID
levels(dd[,47])=c("Pagado totalmente por usted", "Pagado en parte por
usted y en parte por otra persona o instituci?n", "Pagado totalmente por
otra persona o instituci?n", "La actividad era gratuita", "No sabe",
"Negativa", "No aplicable")

#FEDPAIDBY_
levels(dd[,48])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,49])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,50])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,51])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#FEDUSEA
levels(dd[,52])=c("Con mucha frecuencia", "Con cierta frecuencia ",
"Con muy poca frecuencia ", "Con ninguna frecuencia", "Negativa /No
sabe", "No aplicable")

levels(dd[,53])=c("Con mucha frecuencia", "Con cierta frecuencia ",
"Con muy poca frecuencia ", "Con ninguna frecuencia", "Negativa /No
sabe", "No aplicable")

#FEDOUTCOME
levels(dd[,54])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,55])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,56])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,57])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,58])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,59])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

levels(dd[,60])=c("S?", "No", "Negativa/ No sabe", "No aplicable")
####

#NFGUIDEFJT
levels(dd[,61])=c("S?", "No")

#NFEWORKSHOP
levels(dd[,62])=c("S?", "No")

#NFECOURSE

```

```

levels(dd[,63])=c("S?", "No")

#NFELESSON
levels(dd[,64])=c("S?", "No")

#NFEACT01_TYPE
levels(dd[,66])=c("Curso te?rico, charla o conferencia", "Curso, taller
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para
su puesto de trabajo ", "Clases particulares", "No aplicable")

#NFEACT01_PURP
levels(dd[,67])=c("Motivos relacionado con el trabajo", "Motivos personales
o no relacionados con el trabajo", "Negativa/ No sabe", "No aplicable")

#NFEACT01_WORKTIME
levels(dd[,68])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT01_PAIDBY
levels(dd[,69])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT02_TYPE
levels(dd[,70])=c("Curso te?rico, charla o conferencia", "Curso, taller
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para
su puesto de trabajo ", "Clases particulares", "No aplicable")

#NFEACT02_PURP
levels(dd[,71])=c("Motivos relacionado con el trabajo", "Motivos personales
o no relacionados con el trabajo", "Negativa/ No sabe", "No aplicable")

#NFEACT02_WORKTIME
levels(dd[,72])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT02_PAIDBY
levels(dd[,73])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT03_TYPE
levels(dd[,74])=c("Curso te?rico, charla o conferencia", "Curso, taller
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para
su puesto de trabajo ", "Clases particulares", "No aplicable")

#NFEACT03_PURP
levels(dd[,75])=c("Motivos relacionado con el trabajo", "Motivos personales
o no relacionados con el trabajo", "Negativa/ No sabe", "No aplicable")

#NFEACT03_WORKTIME
levels(dd[,76])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT03_PAIDBY
levels(dd[,77])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT04_TYPE

```

```
levels(dd[,78])=c("Curso te?rico, charla o conferencia", "Curso, taller  
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para  
su puesto de trabajo ", "Clases particulares", "No aplicable")
```

```
#NFEACT04_PURP
```

```
levels(dd[,79])=c("Motivos relacionado con el trabajo", "Motivos personales  
o no relacionados con el trabajo", "Negativa/ No sabe", "No aplicable")
```

```
#NFEACT04_WORKTIME
```

```
levels(dd[,80])=c("S?", "No o no trabajaba en ese momento", "Negativa /No  
sabe", "No aplicable")
```

```
#NFEACT04_PAIDBY
```

```
levels(dd[,81])=c("S?", "No o no trabajaba en ese momento", "Negativa /No  
sabe", "No aplicable")
```

```
#NFEACT05_TYPE
```

```
levels(dd[,82])=c("Curso te?rico, charla o conferencia", "Curso, taller  
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para  
su puesto de trabajo ", "Clases particulares", "No aplicable")
```

```
#NFEACT05_PURP
```

```
levels(dd[,83])=c("Motivos relacionado con el trabajo", "Motivos personales  
o no relacionados con el trabajo", "Negativa/ No sabe", "No aplicable")
```

```
#NFEACT05_WORKTIME
```

```
levels(dd[,84])=c("S?", "No o no trabajaba en ese momento", "Negativa /No  
sabe", "No aplicable")
```

```
#NFEACT05_PAIDBY
```

```
levels(dd[,85])=c("S?", "No o no trabajaba en ese momento", "Negativa /No  
sabe", "No aplicable")
```

```
#NFEACT06_TYPE
```

```
levels(dd[,86])=c("Curso te?rico, charla o conferencia", "Curso, taller  
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para  
su puesto de trabajo ", "Clases particulares", "No aplicable")
```

```
#NFEACT06_PURP
```

```
levels(dd[,87])=c("Motivos relacionado con el trabajo", "Motivos personales  
o no relacionados con el trabajo", "Negativa/ No sabe", "No aplicable")
```

```
#NFEACT06_WORKTIME
```

```
levels(dd[,88])=c("S?", "No o no trabajaba en ese momento", "Negativa /No  
sabe", "No aplicable")
```

```
#NFEACT06_PAIDBY
```

```
levels(dd[,89])=c("S?", "No o no trabajaba en ese momento", "Negativa /No  
sabe", "No aplicable")
```

```
#NFEACT07_TYPE
```

```
levels(dd[,90])=c("Curso te?rico, charla o conferencia", "Curso, taller  
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para  
su puesto de trabajo ", "Clases particulares", "No aplicable")
```

```
#NFEACT07_PURP
```

```
levels(dd[,91])=c("Motivos relacionado con el trabajo", "Motivos personales  
o no relacionados con el trabajo", "No aplicable")
```

```

#NFEACT07_WORKTIME
levels(dd[,92])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFEACT07_PAIDBY
levels(dd[,93])=c("S?", "No o no trabajaba en ese momento", "Negativa /No
sabe", "No aplicable")

#NFERAND1 -FALTEN CATEGORIES
levels(dd[,94])

#NFERAND1_TYPE
levels(dd[,95])=c("Curso te?rico, charla o conferencia", "Curso, taller
o seminario que combina teor?a y pr?ctica ", "Actividad planificada para
su puesto de trabajo ", "Clases particulares", "No aplicable")

#NFIELD1
levels(dd[,80])=c("Formaci3n general/b?sica", "Educaci3n", "Artes,
humanidades y lenguas", "Ciencias sociales, periodismo y documentaci3n",
"Negocios, administraci3n y derecho", "Ciencias naturales, qu?micas,
f?sicas y matem?ticas", "Tecnolog?as de la informaci3n y las
comunicaciones (TIC)", "Mec?nica, electr3nica; industria y construcci3n",
"Agricultura, ganader?a, pesca, silvicultura y veterinaria", "Salud y
servicios sociales", "Servicios", "Sectores desconocidos o no
especificados", "Negativa/ No sabe", "No aplicable")

#NFEDIST1
levels(dd[,97])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEDISTOL1
levels(dd[,98])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOERA1
levels(dd[,99])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOERB1
levels(dd[,100])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_01a
levels(dd[,101])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_01b
levels(dd[,102])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_02
levels(dd[,103])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_03
levels(dd[,104])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_04
levels(dd[,105])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_13
levels(dd[,106])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_11

```

```

levels(dd[,107])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_06
levels(dd[,108])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_07
levels(dd[,109])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_08
levels(dd[,110])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_09
levels(dd[,111])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_10
levels(dd[,112])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEREASON1_12
levels(dd[,113])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEPROVIDER1
levels(dd[,115])=c("Una instituci?n de educaci?n oficial (colegios,
institutos, universidades", "Instituciones de educaci?n y formaci?n no
oficial ( por ejemplo academias)", "Instituciones comerciales que no
tienen la educaci?n y la formaci?n como actividad principal (como los
suministradores de equipo) ", "Empleadores ", "Organizaciones
empresariales, c?maras de comercio ", "Sindicatos ", "Asociaciones sin
fines de lucro (fundaciones, sociedades culturales, partidos pol?ticos)",
"Particulares (como estudiantes que dan clases particulares)",
"Instituciones no comerciales que no tienen la educaci?n y formaci?n como
actividad principal (bibliotecas, museos, Administraciones P?blicas)",
"Otros organismos", "Negativa /No sabe", "No aplicable")

#NFECERT1
levels(dd[,116])=c("S?, es exigido por el empleador o requerido
legalmente", "S?, aunque no es exigido por el empleador ni requerido
legalmente", "No (solo se obtiene un certificado de asistencia)",
"Negativa /No sabe", "No aplicable")

#NFEPAID1
levels(dd[,117])=c("Pagado totalmente por usted", "Pagado en parte por
usted y en parte por otra persona o instituci?n", "Pagado totalmente por
otra persona o instituci?n", "La actividad era gratuita", "No sabe",
"Negativa", "No aplicable")

#NFEPAIDBY1_1
levels(dd[,118])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEPAIDBY1_2
levels(dd[,119])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEPAIDBY1_3
levels(dd[,120])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEPAIDBY1_4
levels(dd[,121])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEUSEA1

```

```

levels(dd[,125])=c("Con mucha frecuencia", "Con cierta frecuencia ",
"Con muy poca frecuencia ", "Con ninguna frecuencia", "Negativa /No
sabe", "No aplicable")

#NFEUSEAB1
levels(dd[,126])=c("Con mucha frecuencia", "Con cierta frecuencia ",
"Con muy poca frecuencia ", "Con ninguna frecuencia", "Negativa /No
sabe", "No aplicable")

#NFEOUTCOME1_1
levels(dd[,127])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOUTCOME1_3
levels(dd[,128])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOUTCOME1_2
levels(dd[,129])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOUTCOME1_4
levels(dd[,130])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOUTCOME1_5
levels(dd[,131])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOUTCOME1_6
levels(dd[,132])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NFEOUTCOME1_7
levels(dd[,133])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFICULTY_A
levels(dd[,134])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NEEDA
levels(dd[,135])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFICULTT_B
levels(dd[,136])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#NEEDB
levels(dd[,137])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_01
levels(dd[,138])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_02
levels(dd[,139])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_03
levels(dd[,140])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_04
levels(dd[,141])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_05
levels(dd[,142])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_06

```

```

levels(dd[,143])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_07
levels(dd[,144])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_08a
levels(dd[,145])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_08b
levels(dd[,146])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_10
levels(dd[,147])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_12
levels(dd[,148])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFTYPE_09
levels(dd[,149])=c("S?", "No", "Negativa/ No sabe", "No aplicable")

#DIFFMAIN
levels(dd[,150])=c("No reun?a los requisitos previos", "La formaci?n era demasiado cara o no se lo pod?a permitir", "Falta de apoyo del empleador o de los servicios p?blicos", "La formaci?n no era compatible con el horario de mis tareas o se organiz? en un momento inoportuno", "Las actividades ofrecidas no ten?an lugar a una distancia razonable", "No tenia acceso a un ordenador o Internet ", "No dispon?a de tiempo por responsabilidades familiares", "Por problemas de salud", "Por la edad", "Tuvo dificultades en encontrar lo que quer?a", "Con anterioridad tuvo una mala experiencia en actividades de aprendizaje", "Otras razones personales", "Negativa/ No sabe", "No aplicable")

#INFFAMILY
levels(dd[,151])=c("S?", "No")

#INF MATERIAL
levels(dd[,152])=c("S?", "No")

#INF COMPUTER
levels(dd[,153])=c("S?", "No")

#INF MEDIA
levels(dd[,154])=c("S?", "No")

#INF MUSEUM
levels(dd[,155])=c("S?", "No")

#INF LIBRARIES
levels(dd[,156])=c("S?", "No")

#INF
levels(dd[,157])=c("S?", "No")

#LANGMOTHER1
#levels(dd[,158])

#LANGMOTHER2
#levels(dd[,159])

```



```

#LANGUSED_1
#levels(dd[,161])

#LANGUSED_2
#levels(dd[,162])

#LANGUSED_3
#levels(dd[,163])

#LANGUSED_4
#levels(dd[,164])

#LANGUSED_5
#levels(dd[,165])

#LANGUSED_6
#levels(dd[,166])

#LANGUSED_7
#levels(dd[,167])

#LANGBEST1
#levels(dd[,168])

#LANGBEST2
#levels(dd[,169])

#LANGLEVEL1
levels(dd[,170])=c("Solamente entiende y usa unas pocas palabras y frases",
"Entiende y usa las expresiones m?s comunes. Usa el idioma para las
situaciones y cosas familiares", "Entiende lo esencial del lenguaje
com?n y redacta textos sencillos. Puede describir experiencias y
acontecimientos", "Entiende una gran variedad de textos complicados y usa
el idioma con fluidez. Domina el idioma casi completamente", "Negativa
/No sabe", "No aplicable")

#LANGLEVEL2
levels(dd[,171])=c("Solamente entiende y usa unas pocas palabras y frases",
"Entiende y usa las expresiones m?s comunes. Usa el idioma para las
situaciones y cosas familiares", "Entiende lo esencial del lenguaje
com?n y redacta textos sencillos. Puede describir experiencias y
acontecimientos", "Entiende una gran variedad de textos complicados y usa
el idioma con fluidez. Domina el idioma casi completamente", "Negativa
/No sabe", "No aplicable")

#HHTYPE
levels(dd[,174])=c("Es un hogar unipersonal", "Vive un padre o madre con
alg?n hijo de menos de 25 a?os de edad", "Vive una pareja sola o sin hijos
de menos de 25 a?os de edad", "Vive una pareja con alg?n hijo de menos de
25 a?os de edad ", "Vive una Pareja o un padre o madre con alg?n hijo de
menos de 25 a?os de edad y otras personas ", "Es otro tipo de hogar")

#HHINCOME1
levels(dd[,175])=c("Menos de 650 euros", "De 650 a menos de 900 euros",
"De 900 a menos de 1200 euros", "De 1200 a menos de 1500 euros", "De
1500 a menos de 1800 euros", "De 1800 a menos de 2100 euros", "De
2100 a menos de 2500 euros", "De 2500 a menos de 3000 euros", "De

```

```

3000 a menos de 4000 euros ", "De 4000 euros en adelante", "Negativa/
No sabe")

#INTMETHOD1
levels(dd[,176])=c("Entrevista personal (CAPI) ", "Entrevista telef?nica
(CATI)", "Entrevista por internet (CAWI)", "Entrevista mixta (CAPI y
telef?nica) ")

#DEG_URB
levels(dd[,177])=c("Zona densamente poblada",
                  "Zona intermedia","Zona poco poblada")

#TAMUN
levels(dd[,178])=c("Capital de provincia o con m?s de 100.000
habitantes","Municipios de 50.001 a 100.000 habitantes",
                  "Municipios de 20.001 a 50.000 habitantes",
                  "Municipios de 10.001 a 20.000 habitantes",
                  "Municipios de 10.000 o menos habitantes")

library(car)

HATLEVEL_cat <- dd$HATLEVEL
HATLEVEL_cat <- recode(HATLEVEL_cat, "'Analfabetos '= 'Educaci3n primaria o
inferior' ; 'Estudios primarios incompletos ' = 'Educaci3n primaria o
inferior' ; 'Educaci3n primaria ' = 'Educaci3n primaria o inferior' ;
'Primera etapa de educaci3n secundaria sin t3tulo' = 'Primera etapa de
educaci3n secundaria' ; 'Primera etapa de educaci3n secundaria con t3tulo'
= 'Primera etapa de educaci3n secundaria' ; 'Certificados de
profesionalidad de nivel 1 o nivel 2' = 'Primera etapa de educaci3n
secundaria' ; 'Bachillerato y similares' = 'Segunda etapa de educaci3n
secundaria (bachillerato y similares)' ; 'Enseñanzas de formaci3n
profesional' = 'Segunda etapa de educaci3n secundaria (formaci3n
profesional)' ; 'Formaci3n Profesional B3sica (FPB)' = 'Segunda etapa de
educaci3n secundaria (formaci3n profesional)' ; 'Certificados de
profesionalidad de nivel 3' = 'Segunda etapa de educaci3n secundaria
(formaci3n profesional)' ; 'Enseñanzas de formaci3n profesional / T3tulos
propios universitarios' = 'Formaci3n profesional de grado superior' ;
'Grados universitarios(240 ECTS) / Diplomados universitarios' = 'Grado
universitario (240 cr. ECTS), diplomatura' ; 'Grados universitarios(+240
ECTS) / Licenciados' = 'Grado universitario (+240 cr. ECTS), licenciatura'
; 'M3steres oficiales universitarios' = 'M3ster y Doctorado' ; 'Doctorado
universitario' = 'M3ster y Doctorado' ")

#Reordenamos factores
HATLEVEL_cat <- factor(HATLEVEL_cat, levels =
levels(HATLEVEL_cat)[c(1,6,7,8,2,4,3,5)])

#Añadimos la variable categorizada a la base de datos
dd <- cbind(dd,HATLEVEL_cat)

FEDLEVEL_cat <- dd$FEDLEVEL
FEDLEVEL_cat <- recode(FEDLEVEL_cat, "'Educaci3n b3sica para adultos' =
'Educaci3n primaria o inferior' ; 'M3sica y danza' = 'Educaci3n primaria o
inferior' ; 'E.S.O.(1º-3º)' = 'Primera etapa de educaci3n secundaria' ;
'Secundaria de adultos' = 'Primera etapa de educaci3n secundaria' ;
'Certificados de profesionalidad de nivel 1' = 'Primera etapa de educaci3n
secundaria' ; 'E.S.O (4º)' = 'Primera etapa de educaci3n secundaria' ;
'Bachillerato' = 'Segunda etapa de educaci3n secundaria (Bachillerato y

```

```
similares)' ; 'Formación profesional, artes plásticas, diseño y deportes
grado medio' = 'Segunda etapa de educación secundaria (formación
profesional)' ; 'Enseñanzas profesionales de música y danza' = 'Segunda
etapa de educación secundaria (formación profesional)' ; 'EOI' = 'Segunda
etapa de educación secundaria (formación profesional)' ; 'Cualificación
profesional inicial' = 'Segunda etapa de educación secundaria (formación
profesional)' ; 'Certificados de profesionalidad de nivel 2' = 'Segunda
etapa de educación secundaria (formación profesional)' ; 'Formación
profesional básica' = 'Segunda etapa de educación secundaria (formación
profesional)' ; 'Certificados de profesionalidad de nivel 3' = 'Segunda
etapa de educación secundaria (formación profesional)' ; 'Formación
profesional, artes plásticas, diseño y deportes grado superior' =
'Formación profesional de grado superior' ; 'Títulos propios universitario'
= 'Formación profesional de grado superior' ; 'Grados universitarios (240
ECTS)' = 'Grado universitario (240 cr. ECTS), diplomatura' ; 'Diplomaturas
universitarias' = 'Grado universitario (240 cr. ECTS), diplomatura' ;
'Títulos propios universitarios de experto o especialista' = 'Grado
universitario (240 cr. ECTS), diplomatura' ; 'Grados universitarios (+ 240
ECTS )' = 'Grados universitarios(+240 ECTS) / Licenciados' ; 'Licenciaturas
y equivalentes' = 'Grados universitarios(+240 ECTS) / Licenciados' ;
'Másteres oficiales universitarios' = 'Máster, Ciencias de la Salud y
Doctorado' ; 'Especialidades en Ciencias de la Salud' = 'Máster, Ciencias
de la Salud y Doctorado' ; 'Títulos propios universitarios de máster' =
'Máster, Ciencias de la Salud y Doctorado' ; 'Doctorado ' = 'Máster,
Ciencias de la Salud y Doctorado'")
```

```
FEDLEVEL_cat <- factor(FEDLEVEL_cat, levels =
levels(FEDLEVEL_cat)[c(1,7,8,9,2,3,4,5,6)])
```

```
dd <- cbind(dd,FEDLEVEL_cat)
```

```
#####
# UNIVARIANT SENZILLA EDUCACIO #
#####
```

```
library(ggplot2)
library(plotly)
#setwd("C:/Users/alicia.leon.molina/Documents/TFG")
load("C:/Users/Propietario/Desktop/TFG/dades2.RData")
```

```
v <- c()
```

```
names <- colnames(dd)
```

```
for(i in 1:ncol(dd)){
```

```
  if(class(dd[,i])=="factor"){
```

```
    v <- c(v, i)
```

```
  }
```

```
}
```

```
v <- v[-1]
```

```
#v cont? les posicions de les columnes amb variables categoriques
```

```

x <- c()

for(i in 1:ncol(dd)){

  if(class(dd[,i])=="numeric"){

    x <- c(x, i)

  }
}

#x contiene las posiciones de las columnas con variables numéricas

for(i in v){
  print(names(dd)[i])
  print(summary(dd[,i]))
}

#for(i in v){
#  if(length(levels(dd[,i]))>4){
#    par(las=2)
#    par(mar=c(5,15,4,2))
#    barplot(table(dd[,i]), main=colnames(dd)[i], horiz=T, cex.names=0.8)
#  }else{
#    par(las=1)
#    barplot(table(dd[,i]),main = colnames(dd)[i], cex.names = 0.8)
#  }
#}

for(i in v){
  if(length(levels(dd[,i]))>4){
    par(las=2)
    par(mar=c(5,15,4,2))
    barplot(table(dd$SEX,dd[,i]), main=colnames(dd)[i], horiz=T,
cex.names=0.8, legend.text = c("Hombres","Mujeres"))
  }else{
    par(las=1)
    barplot(table(dd$SEX,dd[,i]),main = colnames(dd)[i], cex.names = 0.8,
legend.text = c("Hombres","Mujeres"))
  }
}

for(i in x){
  print(names(dd)[i])
  print(summary(dd[,i]))
}

for(i in x){
  boxplot(dd[,i]~dd$SEX,main = colnames(dd)[i])
}

#####
# DESCRIPTIVA EDUCACIO #
#####

```

```

d1 <- subset(dd, EDAD_gr=="18-29 a?os")
d2 <- subset(dd, EDAD_gr=="30-40 a?os")
d3 <- subset(dd, EDAD_gr=="41-51 a?os")
d4 <- subset(dd, EDAD_gr=="52-64 a?os")

library(ggplot2)

# SEXO

a <- plyr::count(dd$SEX)
n <- length(dd$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL, fill = "G?nero")

#JOBISCO

s <- ggplot(dd, aes(JOBISCO, fill = SEX)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("G?nero", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Nivel de
estudios",y="Proporci?n")

# MAXIMO NIVEL DE ESTUDIOS (HATLEVEL)

## General x sexo
s <- ggplot(dd, aes(HATLEVEL_cat, fill = SEX)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("G?nero", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Nivel de
estudios",y="Proporci?n")

## General x sexo x edad (ponderado)
gr4 <- ggplot(dd_1, aes(x=HATLEVEL_cat, fill=SEX)) +
geom_bar(position="fill") + scale_fill_manual("G?nero", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Nivel de
Estudios",y="Proporci?n")

```

```

gr4 + coord_flip() + facet_wrap(~ EDAD_gr)

# COTENIDO O TEMA PRINCIPAL DE LOS ESTUDIOS (HATFIELD)

dd_1 <- subset(dd, HATFIELD!="No aplicable")
dd_1 <- subset(dd_1, HATFIELD!="Negativa/ No sabe")

## General x sexo
s <- ggplot(dd_1, aes(HATFIELD, fill = SEX)) + geom_bar(position = "fill")
s + coord_flip() + labs(x="Contenido / Tema",y="Proporci?n") +
theme(plot.title = element_text(hjust =1, vjust = 2, color="gray32")) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

## General x sexo x edad (ponderado)
gr4 <- ggplot(dd_1, aes(x=HATFIELD, fill=SEX)) + geom_bar(position="fill")
+ scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2",
"Mujer" = "darkcyan")) + labs(x="Sector",y="Proporci?n")

gr4 + coord_flip() + facet_wrap(~ EDAD_gr)

# EDAD DE FINALIZACI?N DE LOS ESTUDIOS (HATAGE) TREURE 99

dd_1 <- subset(dd, HATAGE!=88)
d1_1 <- subset(d1, HATAGE!=88)
d2_1 <- subset(d2, HATAGE!=88)
d3_1 <- subset(d3, HATAGE!=88)
d4_1 <- subset(d4, HATAGE!=88)

## General x sexo
s <- ggplot(dd_1, aes(x= HATAGE, color=SEX)) + geom_histogram(fill="white",
alpha=0.5, position = "identity")
s + labs(x="Edad",fill="G?nero") + scale_fill_manual(values =
c("darkgoldenrod2","darkcyan"))

## General x sexo x edad (ponderado)
s <- ggplot(d1_1, aes(x= SEX, y= HATAGE, fill=SEX)) + geom_boxplot()
s + labs(title = "(18-29 a?os)") + scale_fill_manual(values =
c("darkgoldenrod2","darkcyan")) + theme(plot.title = element_text(hjust =
0.5, vjust = 2, color="gray32"), legend.position="none", axis.title =
element_blank())

s <- ggplot(d2_1, aes(x= SEX, y= HATAGE, fill=SEX)) + geom_boxplot()
s + labs(x="G?nero",y="Edad",fill="G?nero",title = "(30-40 a?os)") +
scale_fill_manual(values = c("darkgoldenrod2","darkcyan")) +
theme(plot.title = element_text(hjust = 0.5, vjust = 2, color="gray32"),
legend.position="none", axis.title = element_blank())

s <- ggplot(d3_1, aes(x= SEX, y= HATAGE, fill=SEX)) + geom_boxplot()

```

```

s + labs(x="G?nero",y="Edad",fill="G?nero",title = "(41-51 a?os)") +
scale_fill_manual(values = c("darkgoldenrod2","darkcyan")) +
theme(plot.title = element_text(hjust = 0.5, vjust = 2, color="gray32"),
legend.position="none", axis.title = element_blank())

s <- ggplot(d4_1, aes(x= SEX, y= HATAGE, fill=SEX)) + geom_boxplot()
s + labs(x="G?nero",y="Edad",fill="G?nero",title = "(52-64 a?os)") +
scale_fill_manual(values = c("darkgoldenrod2","darkcyan")) +
theme(plot.title = element_text(hjust = 0.5, vjust = 2, color="gray32"),
legend.position="none", axis.title = element_blank())

# ABANDONO DE ALGUNA OTRA ENSE?ANZA (DROPEDEC)

dd_1 <- subset(dd, DROPEDEC!="Negativa/ No sabe")
dd_1 <- subset(dd_1, DROPEDEC!="No aplicable")
d_a <- subset(dd_1,DROPEDEC=="S?")

d1_a <- subset(d_a, EDAD_gr=="18-29 a?os")
d2_a <- subset(d_a, EDAD_gr=="30-40 a?os")
d3_a <- subset(d_a, EDAD_gr=="41-51 a?os")
d4_a <- subset(d_a, EDAD_gr=="52-64 a?os")
## General

library(dplyr)
a <- plyr::count(dd_1$DROPEDEC)
n <- length(dd_1$DROPEDEC)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white",size=7)+
scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Abandono")

## cuando abandonan, proporcion sex

a <- plyr::count(d_a$SEX)
n <- length(d_a$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

```

```

mycols <- c("darkgoldenrod2","darkcyan")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")

#Per edats

a <- plyr::count(d4_a$SEX)
n <- length(d4_a$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero",title = "(52-64
a?os)")

# ENSE?ANZA ABANDONADA (DROPEUCLEVEL)

dd_1 <- subset(dd, DROPEUCLEVEL!="Negativa/ No sabe")
dd_1 <- subset(dd_1, DROPEUCLEVEL!="No aplicable")
d_a <- subset(dd_1,DROPEUC=="S?")

d1_a <- subset(d_a, EDAD_gr=="18-29 a?os")
d2_a <- subset(d_a, EDAD_gr=="30-40 a?os")
d3_a <- subset(d_a, EDAD_gr=="41-51 a?os")
d4_a <- subset(d_a, EDAD_gr=="52-64 a?os")

## General
p <-
ggplot(dd_1,aes(x=DROPEUCLEVEL))+geom_bar(width=1,fill="darkgoldenrod2",ae
s(y=(..count..)/sum(..count..)))

```



```

p + coord_flip() + labs(x="",y="Frecuencia",title="Abandono de los
estudios") + theme(plot.title = element_text(hjust = 0.5, vjust = 2,
color="gray32"))

## General x sexo
s <- ggplot(dd_1, aes(DROPEDUCLEVEL, fill = SEX)) + geom_bar(position =
"fill")
s + coord_flip() + labs(x="",y="Proporci?n",title="Abandono de los estudios
seg?n g?nero") + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("G?nero", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan"))

# ACTIVIDAD EDUCATIVA OFICIAL EN LOS ULTIMOS 12 MESES (FEDNUM)

d_a <- subset(dd,FEDNUM>0)

## General x sexo
ggplot(d_a, aes(FEDNUM, fill=SEX)) + geom_bar(aes(FEDNUM,fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + xlab("N?mero de
actividades")+ylab("Proporci?n") + ylim(c(0,0.5)) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

# FEDLEVEL

dd_1 <- subset(dd, FEDLEVEL_cat != "No aplicable")

s <- ggplot(dd_1, aes(FEDLEVEL_cat, fill = SEX)) + geom_bar(position =
"fill")
s + coord_flip() + labs(x="Nivel de estudios",y="Proporci?n") +
theme(plot.title = element_text(hjust =1, vjust = 2, color="gray32")) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

# FEDFIELD

dd_1 <- subset(dd, FEDFIELD != "No aplicable")
dd_1 <- subset(dd_1, FEDFIELD != "Negativa/ No sabe")

s <- ggplot(dd_1, aes(FEDFIELD, fill = SEX)) + geom_bar(position = "fill")
s + coord_flip() + labs(x="Contenido / Tema",y="Proporci?n") +
theme(plot.title = element_text(hjust =1, vjust = 2, color="gray32")) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

# LA EDUCACION EN LOS ULTIMOS 12 MESES ERA A DISTANCIA (FEDDIST)

dd_1 <- subset(dd, FEDDIST!="Negativa/ No sabe")
dd_1 <- subset(dd_1, FEDDIST!="No aplicable")
d_a <- subset(dd_1,FEDDIST=="S?")

## General

```

```

library(dplyr)
a <- plyr::count(dd_1$FEDDIST)
n <- length(dd_1$FEDDIST)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white",size=7)+
scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Educaci?n a distancia")

## si son a distancia, proporcion sex

a <- plyr::count(d_a$SEX)
n <- length(d_a$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")
# Razones para realizar los estudios (FEDREASON)

library(reshape)
library(readxl)
fedreason <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Document/fedreason.xlsx"
)
fedreason<-fedreason[,c(-2,-3,-4)]
fedreason<-as.data.frame(fedreason)
fedreason_data <- melt(data=fedreason, id.vars="razon")
colnames(fedreason_data)[2]<-"G?nero"
levels(fedreason_data$G?nero)[1]<-"Hombre"

```

```

levels(fedreason_data$G?nero)[2]<-"Mujer"

s<-ggplot(fedreason_data, aes(x=razon, y=value, fill = G?nero)) +
  geom_bar(position = "fill", stat = "identity",width=0.9) + coord_flip() +
  scale_fill_manual(values=c("darkgoldenrod2","darkcyan"))
s + aes(razon,fill=G?nero) + xlab("Raz?n") + ylab("Proporci?n")

# ACTIVIDAD REALIZADA EN HORAS DE TRABAJO (FEDWORKTIME)

dd_1 <- subset(dd, FEDWORKTIME!="No aplicable")
dd_1 <- subset(dd_1, FEDWORKTIME!="Negativa /No sabe")
d1_1 <- subset(d1, FEDWORKTIME!="No aplicable")
d1_1 <- subset(d1_1, FEDWORKTIME!="Negativa /No sabe")
d2_1 <- subset(d2, FEDWORKTIME!="No aplicable")
d2_1 <- subset(d2_1, FEDWORKTIME!="Negativa /No sabe")
d3_1 <- subset(d3, FEDWORKTIME!="No aplicable")
d3_1 <- subset(d3_1, FEDWORKTIME!="Negativa /No sabe")
d4_1 <- subset(d4, FEDWORKTIME!="No aplicable")
d4_1 <- subset(d4_1, FEDWORKTIME!="Negativa /No sabe")

## General
p <-
ggplot(dd, aes(x=FEDWORKTIME))+geom_bar(width=1,fill="darkgoldenrod2",aes(y=
(..count..)/sum(..count..)))
p + labs(x="",y="Frecuencia",title="Actividad realizada en horas de
trabajo") + theme(plot.title = element_text(hjust = 0.5, vjust = 2,
color="gray32"))

# PAGO DE LOS GASTOS (FEDPAID)

dd_1 <- subset(dd, FEDPAID!="No aplicable")
dd_1 <- subset(dd_1, FEDPAID!="Negativa")
dd_1 <- subset(dd_1, FEDPAID!="No sabe")

#General x Sexo
ggplot(dd_1, aes(FEDPAID, fill=SEX)) +
geom_bar(aes(stringr::str_wrap(FEDPAID, 30),fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + coord_flip()+
xlab("Pago")+ylab("Proporci?n") + ylim(c(0,0.3)) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

# USO DE LOS CONOCIMIENTOS ADQUIRIDOS (FEDUSEA)

dd_1 <- subset(dd, FEDUSEA!="No aplicable")
dd_1 <- subset(dd_1, FEDUSEA!="Negativa /No sabe")

ggplot(dd_1, aes(FEDUSEA, fill=SEX)) +
geom_bar(aes(stringr::str_wrap(FEDUSEA, 30),fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + coord_flip()+
xlab("Frecuencia")+ylab("Proporci?n") + ylim(c(0,0.3)) +

```

```
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))
```

```
# OBJETIVOS CONSEGUIDOS GRACIAS A LAS FORMACIONES (FEDOUTCOME)
```

```
library(reshape)
```

```
fedoutcome<-as.data.frame(fedoutcome)
fedoutcome_data <- melt(data=fedoutcome, id.vars="b")
colnames(fedoutcome_data)[2]<-"G?nero"
```

```
s<-ggplot(fedoutcome_data, aes(x=b, y=value, fill = G?nero)) +
  geom_bar(position = "fill", stat = "identity",width=0.9) + coord_flip() +
scale_fill_manual(values=c("darkgoldenrod2","darkcyan"))
s + aes(stringr::str_wrap(b, 50),fill=G?nero) + xlab("Objetivos") +
ylab("Proporci?n")
```

```
# EDUCACION NO FORMAL puesto de trabajo (NFEGUIDEDJT)
```

```
d_a <- subset(dd,NFEGUIDEDJT=="S?")
```

```
## General
```

```
library(dplyr)
a <- plyr::count(dd$NFEGUIDEDJT)
n <- length(dd$NFEGUIDEDJT)
```

```
a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))
```

```
count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)
```

```
mycols <- c("paleturquoise3","salmon1")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)
```

```
pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Formaci?n")
```

```
## cuando abandonan, proporcion sex
```

```
a <- plyr::count(d_a$SEX)
n <- length(d_a$SEX)
```

```
a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))
```

```
count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)
```

```
mycols <- c("darkgoldenrod2","darkcyan")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)
```

```
pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")
```

```
#CURSO O SEMINARIO (NFEWORKSHOP)
```

```
d_a <- subset(dd,NFEWORKSHOP=="S?")
```

```
## General
```

```
library(dplyr)
a <- plyr::count(dd$NFEWORKSHOP)
n <- length(dd$NFEWORKSHOP)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))
```

```
count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)
```

```
mycols <- c("paleturquoise3","salmon1")
```

```
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size = 7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)
```

```
pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Formaci?n")
```

```
## cuando abandonan, proporcion sex
```

```
a <- plyr::count(d_a$SEX)
```

```
n <- length(d_a$SEX)
```

```
a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))
```

```

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size = 7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")

#CURSO TEORICO O CHARLA (NFECOURSE)
d_a <- subset(dd,NFECOURSE=="S?")

## General

library(dplyr)
a <- plyr::count(dd$NFECOURSE)
n <- length(dd$NFECOURSE)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Curso")

## cuando abandonan, proporcion sex

a <- plyr::count(d_a$SEX)
n <- length(d_a$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")

```

```

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                              axis.text = element_blank(),
                              axis.ticks = element_blank(),
                              plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL, fill = "G?nero")

#NUMERO TOTAL DE ACTIVIDADES DE EDUCACION NO FORMAL (NFENUM)

d_a <- subset(dd,NFENUM>0)
d_a <- subset(d_a,NFENUM<10)
d_a$NFENUM <- as.factor(d_a$NFENUM)
## General x sexo
ggplot(d_a, aes(NFENUM, fill=SEX)) + geom_bar(aes(NFENUM,fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + xlab("N?mero de
actividades")+ylab("Proporci?n") + ylim(c(0,0.30)) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

#RAZON (NFEACT01_PURP)

library(reshape)
library(readxl)

nfeact_purp <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Educacion/nfeact_purp.xl
sx")
nfeact_purp<-as.data.frame(nfeact_purp)
nfeact_purp_data <- melt(data=nfeact_purp, id.vars="motivo")
colnames(nfeact_purp_data)[2]<-"Genero"
levels(nfeact_purp_data$G?nero)[1]<-"Hombre"
levels(nfeact_purp_data$G?nero)[2]<-"Mujer"

s<-ggplot(nfeact_purp_data, aes(x=motivo,y= value, fill = G?nero)) +
geom_bar(stat="identity", position=position_dodge()) +
scale_fill_manual(values=c("darkgoldenrod2","darkcyan")) + ylim(c(0,0.4))
s + xlab("Motivo") + ylab("Proporci?n")

#REALIZADA EN HORAS DE TRABAJO (NFEACT01_WORKTIME)

library(reshape)
library(readxl)

nfeact_worktime <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Educacion/nfeact_worktim
e.xlsx")
nfeact_worktime<-as.data.frame(nfeact_worktime)
nfeact_worktime_data <- melt(data=nfeact_worktime, id.vars="Trabajo")

```

```

colnames(nfeact_worktime_data)[2]<-"G?nero"
levels(nfeact_worktime_data$G?nero)[1]<-"Hombre"
levels(nfeact_worktime_data$G?nero)[2]<-"Mujer"
nfeact_worktime_data$Trabajo <- as.factor(nfeact_worktime_data$Trabajo)
levels(nfeact_worktime_data) <- c("S?", "No o no trabajaba en ese momento")

s<-ggplot(nfeact_worktime_data, aes(x=reorder(Trabajo, value), y= value,
fill = G?nero)) + geom_bar(stat="identity", position=position_dodge()) +
scale_fill_manual(values=c("darkgoldenrod2", "darkcyan")) + ylim(c(0,0.4))
s + xlab("Actividad realizada en horas de trabajo") + ylab("Proporci?n")

#PAGADA POR SU EMPLEADOR (NFEACT01_PAIDBY)

library(reshape)
library(readxl)

nfeact_paidby <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Educacion/nfeact_paidby.
xlsx")
nfeact_paidby<-as.data.frame(nfeact_paidby)
nfeact_paidby_data <- melt(data=nfeact_paidby, id.vars="Pago")
colnames(nfeact_paidby_data)[2]<-"G?nero"
levels(nfeact_paidby_data$G?nero)[1]<-"Hombre"
levels(nfeact_paidby_data$G?nero)[2]<-"Mujer"
nfeact_paidby_data$Pago <- as.factor(nfeact_paidby_data$Pago)
levels(nfeact_paidby_data) <- c("S?", "No o no trabajaba en ese momento")

s<-ggplot(nfeact_paidby_data, aes(x=reorder(Pago, -value), y= value, fill =
G?nero)) + geom_bar(stat="identity", position=position_dodge()) +
scale_fill_manual(values=c("darkgoldenrod2", "darkcyan")) + ylim(c(0,0.3))
s + xlab("Actividad pagada por su empelador") + ylab("Proporci?n")

#RAZONES PARA PARTICIPAR (NFEREASON1)

library(reshape)
library(readxl)

nfereason <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Educacion/NFEREASON.xlsx
")
nfereason<-as.data.frame(nfereason)
nfereason_data <- melt(data=nfereason, id.vars="razon")
colnames(nfereason_data)[2]<-"Genero"
levels(nfereason_data$Genero)[1]<-"Hombre"
levels(nfereason_data$Genero)[2]<-"Mujer"

s<-ggplot(nfereason_data, aes(x=reorder(razon,value), y=value, fill =
Genero)) +
  geom_bar(position = "fill", stat = "identity", width=0.9) + coord_flip() +
scale_fill_manual(values=c("darkgoldenrod2", "darkcyan"))
s + aes(razon, fill=Genero) + xlab("Raz?n") + ylab("Proporci?n")

#QUIEN ORGANIZA ACTIVIDAD (NFEPROVIDER1)

```



```

dd_1 <- subset(dd, NFEPROVIDER1!="No aplicable")
dd_1 <- subset(dd_1, NFEPROVIDER1!="Negativa /No sabe")

s <- ggplot(dd_1, aes(NFEPROVIDER1, fill = SEX)) + geom_bar(position =
"fill")
s + coord_flip() + labs(x="Horario de realizaci?n",y="Proporci?n") +
theme(plot.title = element_text(hjust =0.5, vjust = 2, color="gray32")) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))+ aes(stringr::str_wrap(NFEPROVIDER1, 50),fill=SEX) +
xlab("Organizaci?n") + ylab("Proporci?n")

#QUIEN PAGA (NFEPAID1)

dd_1 <- subset(dd, NFEPAID1!="No aplicable")
dd_1 <- subset(dd_1, NFEPAID1!="Negativa")
dd_1 <- subset(dd_1, NFEPAID1!="No sabe")

#General x Sexo
ggplot(dd_1, aes(NFEPAID1, fill=SEX)) + geom_bar(aes(NFEPAID1,fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + xlab("Pago de la
actividad")+ylab("Proporci?n")+ scale_fill_manual("G?nero", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan")) + ylim(c(0,0.3)) +
coord_flip()

#QUIEN PAGA (NFEPAIDBY1...) juntar

## General

## General x sexo

## General x sexo x edad (ponderado)

#OBJETIVOS CONSEGUIDOS (NFEOUTCOME1) juntar

## General

## General x sexo

## General x sexo x edad (ponderado)

#HA PARTICIPADO EN EDUCACI?N ULTIMOS 12 MESES Y QUISO PARTICIPAR EN OTRA
ACT(DIFFICULTY_A)

dd_1 <- subset(dd, DIFFICULTY_A!="No aplicable")
dd_1 <- subset(dd_1, DIFFICULTY_A!="Negativa/ No sabe")

d_a <- subset(dd_1,DIFFICULTY_A=="S?")

```

```

## General

library(dplyr)
a <- plyr::count(dd_1$DIFFICULTY_A)
n <- length(dd_1$DIFFICULTY_A)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Formaci?n")

## cuando dicen s?, proporcion sex

a <- plyr::count(d_a$SEX)
n <- length(d_a$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")

#SI DICE NO - CONSIDERA SUFICIENTE LA EDUCACI?N (NEEDA)

dd_1 <- subset(dd, NEEDA!="No aplicable")
dd_1 <- subset(dd_1, NEEDA!="Negativa/ No sabe")

```

```

d_a <- subset(dd_1,NEEDA=="S?")
d_n <- subset(dd_1,NEEDA=="No")

## General

library(dplyr)
a <- plyr::count(dd_1$NEEDA)
n <- length(dd_1$NEEDA)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Formaci?n")

## cuando abandonan, proporcion sex

a <- plyr::count(d_n$SEX)
n <- length(d_n$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")

#NO HA PARTICIPADO EN EDUCACI?N ULTIMOS 12 MESES PERO QUISO (DIFFICULTY_B)

dd_1 <- subset(dd, DIFFICULTY_B!="No aplicable")
dd_1 <- subset(dd_1, DIFFICULTY_B!="Negativa/ No sabe")

```

```

d_a <- subset(dd_1,DIFFICULTY_B=="S?")

## General

library(dplyr)
a <- plyr::count(dd_1$DIFFICULTY_B)
n <- length(dd_1$DIFFICULTY_B)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Formaci?n")

## cuando abandonan, proporcion sex

a <- plyr::count(d_a$SEX)
n <- length(d_a$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")

#SI DICE NO - CONSIDERA SUFICIENTE EDUCACI?N (NEEDB)

```

```

dd_1 <- subset(dd, NEEDB!="No aplicable")
dd_1 <- subset(dd_1, NEEDB!="Negativa/ No sabe")

d_a <- subset(dd_1,NEEDB=="S?")
d_n <- subset(dd_1,NEEDB=="No")

## General

library(dplyr)
a <- plyr::count(dd_1$NEEDB)
n <- length(dd_1$NEEDB)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("paleturquoise3","salmon1")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white",size=7)+
scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "Formaci?n")

## cuando abandonan, proporcion sex

a <- plyr::count(d_n$SEX)

n <- length(d_n$SEX)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")

pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
axis.text = element_blank(),
axis.ticks = element_blank(),
plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL,fill = "G?nero")

```

```

#MOTIVO MAS IMPORTANTE (DIFFMAIN)

dd_1 <- subset(dd, DIFFMAIN!="No aplicable")
dd_1 <- subset(dd_1, DIFFMAIN!="Negativa/ No sabe")

s <- ggplot(dd_1, aes(DIFFMAIN, fill = SEX)) + geom_bar(position = "fill")
s + coord_flip() + labs(x="Horario de realizaci?n",y="Proporci?n") +
theme(plot.title = element_text(hjust =0.5, vjust = 2, color="gray32")) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))+ aes(stringr::str_wrap(DIFFMAIN, 50),fill=SEX) +
xlab("Raz?n") + ylab("Proporci?n")

#APRENDIZAJE AUTONOMO (INF)

ggplot(dd, aes(INF, fill=SEX)) + geom_bar(aes(INF,fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + xlab("Aprendizaje
deliberado")+ylab("Proporci?n") + ylim(c(0,0.30)) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

#IDIOMAS

dd_1 <- subset(dd, LANGUSED<6)
dd_1$LANGUSED <- as.factor(dd_1$LANGUSED)

ggplot(dd_1, aes(LANGUSED, fill=SEX)) + geom_bar(aes(LANGUSED,fill=SEX,
(..count..)/sum(..count..)),position = "dodge") + xlab("N?
idiomas")+ylab("Proporci?n") + ylim(c(0,0.30)) +
scale_fill_manual("G?nero", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

#####
### Test Chi-cuadrado ###
#####

# Nivel de estudios

datos <- data.frame(sexo = dd$SEX, estudios = dd$HATLEVEL_cat)

```

```

tabla <- table(datos$sexo, datos$estudios, dnn=c("Género", "Nivel de
estudios"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Contenido de los estudios

dd_1 <- subset(dd, HATFIELD!="No aplicable")
dd_1 <- subset(dd_1, HATFIELD!="Negativa/ No sabe")

datos <- data.frame(sexo = dd_1$SEX, estudios = dd_1$HATFIELD)

tabla <- table(datos$sexo, datos$estudios, dnn=c("Género", "Nivel de
estudios"))
tabla <- tabla[,-13]
tabla <- tabla[,-13]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Edad de finalización de los estudios

dd_1 <- subset(dd, HATAGE!=88)

t.test(HATAGE~SEX,data=dd_1,alternative="two.sided",conf.level=.95)

d1 <- subset(dd_1, EDAD_gr=="18-29 años")
d2 <- subset(dd_1, EDAD_gr=="30-40 años")
d3 <- subset(dd_1, EDAD_gr=="41-51 años")
d4 <- subset(dd_1, EDAD_gr=="52-64 años")

t.test(HATAGE~SEX,data=d1,alternative="two.sided",conf.level=.95)
t.test(HATAGE~SEX,data=d2,alternative="two.sided",conf.level=.95)
t.test(HATAGE~SEX,data=d3,alternative="two.sided",conf.level=.95)
t.test(HATAGE~SEX,data=d4,alternative="two.sided",conf.level=.95)

# Abandono de alguna otra enseñanza

dd_1 <- subset(dd, DROPEduc!="Negativa/ No sabe")
dd_1 <- subset(dd_1, DROPEduc!="No aplicable")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$DROPEduc)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Abandono
educación"))
tabla <- tabla[,-3]
tabla <- tabla[,-3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Actividades realizadas en los últimos 12 meses

```

```

d_a <- subset(dd, FEDNUM>0)

t.test(FEDNUM~SEX, data=d_a, alternative="two.sided", conf.level=.95)

# Nivel de estudios estudiantes actuales

dd_1 <- subset(dd, FEDLEVEL_cat != "No aplicable")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDLEVEL_cat)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Nivel de estudios"))
tabla <- tabla[, -9]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Contenido de los estudios actuales

dd_1 <- subset(dd, FEDFIELD != "No aplicable")
dd_1 <- subset(dd_1, FEDFIELD != "Negativa/ No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDFIELD)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Contenido de los
estudios"))
tabla <- tabla[, -13]
tabla <- tabla[, -13]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Educación a distancia

dd_1 <- subset(dd, FEDDIST!="Negativa/ No sabe")
dd_1 <- subset(dd_1, FEDDIST!="No aplicable")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDDIST)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Educación a
distancia"))
tabla <- tabla[, -3]
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Actividad realizada en horas de trabajo remuneradas

dd_1 <- subset(dd, FEDWORKTIME!="No aplicable")
dd_1 <- subset(dd_1, FEDWORKTIME!="Negativa /No sabe")

```



```

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDWORKTIME)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Actividad en horas
de trabajo"))
tabla <- tabla[, -6]
tabla <- tabla[, -6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Pago de los gastos de la actividad

dd_1 <- subset(dd, FEDPAID!="No aplicable")
dd_1 <- subset(dd_1, FEDPAID!="Negativa")
dd_1 <- subset(dd_1, FEDPAID!="No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDPAID)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Pago de la
actividad"))
tabla <- tabla[, -5]
tabla <- tabla[, -5]
tabla <- tabla[, -5]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Frecuencia de uso de los conocimientos adquiridos

dd_1 <- subset(dd, FEDUSEA!="No aplicable")
dd_1 <- subset(dd_1, FEDUSEA!="Negativa /No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDUSEA)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Frecuencia uso
conocimientos"))
tabla <- tabla[, -5]
tabla <- tabla[, -5]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Objetivos a los que ha contribuido la actividad

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$DROPEduc)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Abandono
educación"))
tabla <- tabla[, -3]
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

```

```

# Dedicación a estudiar en horas de trabajo

dd_1 <- subset(dd, FEDWORKTIME!="No aplicable")
dd_1 <- subset(dd_1, FEDWORKTIME!="Negativa /No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$FEDWORKTIME)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Estudio en horas de
trabajo"))
tabla <- tabla[, -7]
tabla <- tabla[, -6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Participación en educación no formal

datos <- data.frame(sexo = dd$SEX, var = dd$NFEGUIDEDJT)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Formación"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Participación en cursos

datos <- data.frame(sexo = dd$SEX, var = dd$NFEWORKSHOP)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Cursos"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Participación en charlas

datos <- data.frame(sexo = dd$SEX, var = dd$NFE COURSE)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Charlas"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Actividades no formales en total

d_a <- subset(dd, NFENUM>0)
d_a <- subset(d_a, NFENUM<10)

t.test(NFENUM~SEX, data=d_a, alternative="two.sided", conf.level=.95)

# Razón para realizar una actividad no formal

```

```

mat <- matrix(c(6081,1514,6037,2220),ncol = 2)
colnames(mat) <- c("Hombre","Mujer")
rownames(mat) <- c("Motivos relacionado con el trabajo","Motivos personales
o no relacionados con el trabajo")

tabla <- as.table(mat)

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Actividad realizada en horas de trabajo

mat <- matrix(c(3526,3888,2868,5242),ncol = 2)
colnames(mat) <- c("Hombre","Mujer")
rownames(mat) <- c("Sí","No o no trabajaba en ese momento")

tabla <- as.table(mat)

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

#PAGADA POR SU EMPLEADOR (NFEACT01_PAIDBY)

mat <- matrix(c(4103,3170,3790,4127),ncol = 2)
colnames(mat) <- c("Hombre","Mujer")
rownames(mat) <- c("Sí","No o no trabajaba en ese momento")

tabla <- as.table(mat)

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Intención de participar en otras actividades extra

dd_1 <- subset(dd, DIFFICULTY_A!="No aplicable")
dd_1 <- subset(dd_1, DIFFICULTY_A!="Negativa/ No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$DIFFICULTY_A)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Inteción"))
tabla <- tabla[,-3]
tabla <- tabla[,-3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Considera suficiente los estudios cursados

```

```

dd_1 <- subset(dd, NEEDA!="No aplicable")
dd_1 <- subset(dd_1, NEEDA!="Negativa/ No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$NEEDA)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Suficiente"))
tabla <- tabla[, -3]
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

#Intencion de participar en alguna educación

dd_1 <- subset(dd, DIFFICULTY_B!="No aplicable")
dd_1 <- subset(dd_1, DIFFICULTY_B!="Negativa/ No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$DIFFICULTY_B)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Inteción"))
tabla <- tabla[, -3]
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Motivos para no participar en educación

dd_1 <- subset(dd, DIFFMAIN!="No aplicable")
dd_1 <- subset(dd_1, DIFFMAIN!="Negativa/ No sabe")

datos <- data.frame(sexo = dd_1$SEX, var = dd_1$DIFFMAIN)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Motivos"))
tabla <- tabla[, -13]
tabla <- tabla[, -13]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Aprendizaje deliberado

datos <- data.frame(sexo = dd$SEX, var = dd$INF)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Aprendizaje"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Numero de idiomas diferente al materno

dd_1 <- subset(dd, LANGUSED<6)

```

```
t.test(LANGUSED~SEX,data=dd,alternative="two.sided",conf.level=.95)
t.test(LANGUSED~SEX,data=dd_1,alternative="two.sided",conf.level=.95)
```

```
#####
# CLUSTERING EDUCACIO #
#####
```

```
library(cluster)
library(dendextend)
library(FactoMineR)
library(dplyr)
library(factoextra)
library(ggpubr)
```

```
# ESTUDIANTS GENERAL #
#####
```

```
load("C:/Users/34689/Documents/Tfg ali/Scripts/educacio/dades5.RData")
```

```
dades_clust1 <- subset(dd,select=c(SEX, HATLEVEL_cat, HATFIELD, DROPEduc))
dades_clust1 <- subset(dades_clust1, HATFIELD!="Negativa/ No sabe" &
HATFIELD!="No aplicable")
```

```
dissimMatrix <- daisy(dades_clust1, metric = "gower", stand=TRUE)
```

```
distMatrix<-dissimMatrix^2
```

```
h1 <- hclust(distMatrix,method="ward.D")
```

```
plot(h1)
```

```
c2 <- cutree(h1,2)
table(c2)
```

```
fviz_dend(x=h1, k=2, cex=0.6, k_colors = c("darkgoldenrod2","darkcyan")) +
geom_hline(yintercept = 1500, linetype = "dashed")
```

```
dades_clust1 <- mutate(dades_clust1, cluster = c2)
dades_clust1$cluster <- as.factor(dades_clust1$cluster)
levels(dades_clust1$cluster) <- c("Cl?ster 1", "Cl?ster 2")
```

```
by(dades_clust1,dades_clust1$cluster,summary)
```

```
catdes(dades_clust1,5)
```

```
###REPRESENTACI? PER ACP
```

```
#pca_dades1 <- prcomp(dades_clust1)
#p <- fviz_pca_ind(X=pca_dades1, habillage = dades_clust1$SEX, geom =
"point") +
```

```

#       theme_bw() + theme(legend.position = "bottom")
#p

#fviz_cluster(object = list(data=dades_clust1, cluster=cutree(h1, k=2)),
#               ellipse.type = "convex", repel = TRUE, show.clust.cent =
FALSE,
#               labels.size = 8) +
#   labs(title = "Hierarchical clustering + Proyecci?n PCA",
#         subtitle = "Distancia eucl?dea, Lincage complete, K=2") +
#   theme_bw() +
#   theme(legend.position = "bottom")

# SENSE VARIABLE SEXE

dissimMatrix2 <- daisy(dades_clust1[, -1], metric = "gower", stand=TRUE)
distMatrix2 <- dissimMatrix2^2

h2 <- hclust(distMatrix2, method="ward.D")
plot(h2)

## Dos clusters
c2 <- cutree(h2, 2)

#dendograma <- as.dendrogram(h1)
#dend_1500 <- color_branches(dendograma, h = 1500)
#plot(dend_1500)

fviz_dend(x=h2, k=2, cex=0.6) + geom_hline(yintercept = 1500, linetype =
"dashed")

dades_clust1 <- mutate(dades_clust1, cluster = c2)
dades_clust1$cluster <- as.factor(dades_clust1$cluster)

by(dades_clust1, dades_clust1$cluster, summary)

catdes(dades_clust1, 5)

## Tres clusters

c3 <- cutree(h2, 3)

dendograma <- as.dendrogram(h2)
dend_900 <- color_branches(dendograma, h = 900)
plot(dend_900)

#fviz_dend(x=h2, k=3, cex=0.6) + geom_hline(yintercept = 900, linetype =
"dashed")

dades_clust1 <- mutate(dades_clust1, cluster = c3)
dades_clust1$cluster <- as.factor(dades_clust1$cluster)

by(dades_clust1, dades_clust1$cluster, summary)

catdes(dades_clust1, 5)

#COMPARATIVA CLUSTERS
dend_1 <- as.dendrogram(h1)

```

```

dend_2 <- as.dendrogram(h2)
tanglegram(dend1 = dend_1, dend2 = dend_2, highlight_distinct_edges = TRUE,
            common_subtrees_color_branches = TRUE)

## ESTUDIANTS ACTUALS ##
#####

dades_clust2 <- subset(dd, select = c(SEX, FEDNUM, FEDLEVEL_cat, FEDFIELD,
FEDWORKTIME, FEDPAID))
dades_clust2 <- subset(dades_clust2, FEDLEVEL_cat!="No aplicable")

#dissimilarity matrix

dissimMatrix <- daisy(dades_clust2, metric = "gower", stand=TRUE)

distMatrix<-dissimMatrix^2

h1 <- hclust(distMatrix,method="ward.D")

c2 <- cutree(h1,2)
table(c2)

fviz_dend(x=h1, k=2, cex=0.6, k_colors = c("darkcyan","darkgoldenrod2")) +
geom_hline(yintercept = 200, linetype = "dashed")

dades_clust2 <- mutate(dades_clust2, cluster = c2)
dades_clust2$cluster <- as.factor(dades_clust2$cluster)

by(dades_clust2,dades_clust2$cluster,summary)

catdes(dades_clust2,7)

# SENSE VARIABLE SEXE

dissimMatrix <- daisy(dades_clust2[, -1], metric = "gower", stand=TRUE)

distMatrix<-dissimMatrix^2

h1 <- hclust(distMatrix,method="ward.D")

plot(h1, main = "")

c2 <- cutree(h1,3)
table(c2)

fviz_dend(x=h1, k=3, cex=0.6) + geom_hline(yintercept = 100, linetype =
"dashed")

dades_clust2 <- mutate(dades_clust2, cluster = c2)
dades_clust2$cluster <- as.factor(dades_clust2$cluster)

```

```

by(dades_clust2,dades_clust2$cluster,summary)

catdes(dades_clust2,7)

#####
# TRACTAMENT DE DADES INSERCIO #
#####

dd <- fichero_salida
rm(fichero_salida)

# SEXO

dd$SEXO <- as.factor(dd$SEXO)
levels(dd$SEXO) = c("Hombre","Mujer")
summary(dd$SEXO)

# RAMA

dd$RAMA <- as.factor(dd$RAMA)
levels(dd$RAMA) = c("Artes y humanidades","Ciencias","Ciencias sociales y
jurídicas","Ingeniería y arquitectura","Ciencias de la salud")

# AMBITO DE ESTUDIO

dd$AMBITO <- as.factor(dd$AMBITO)
levels(dd$AMBITO) = c("Formación de docentes de enseñanza infantil",
"Formación de docentes de enseñanza primaria", "Otra formación de
personal docente y ciencias de la educación", "Artes", "Técnicas
audiovisuales y medios de comunicación", "Humanidades", "Lenguas",
"Psicología", "Economía", "Otras ciencias sociales y del
comportamiento", "Periodismo e información", "Administración y gestión
de empresas", "Otra educación comercial y empresarial", "Derecho",
"Ciencias de la vida", "Ciencias físicas, químicas, geológicas",
"Matemáticas y estadística", "Informática", "Ingeniería y profesiones
afines", "Industria manufacturera y producción", "Arquitectura y
construcción", "Agricultura, ganadería y pesca", "Veterinaria",
"Medicina", "Enfermería y atención a enfermos", "Otras ciencias de la
salud", "Trabajo social y orientación", "Turismo y hostelería",
"Deportes", "Servicios de transporte")

# VARIABLE NUEVA SECTOR

df <- subset(dd, dd$AMBITO=="Lenguas" | dd$AMBITO=="Humanidades" |
dd$AMBITO=="Artes" | dd$AMBITO=="Ciencias de la vida" |
dd$AMBITO=="Ciencias físicas, químicas, geológicas" | dd$AMBITO=="Formación
de docentes de enseñanza infantil" | dd$AMBITO=="Formación de docentes de
enseñanza primaria" | dd$AMBITO=="Otra formación de personal docente y
ciencias de la educación" | dd$AMBITO=="Administración y gestión de
empresas" | dd$AMBITO=="Otra educación comercial y empresarial" |
dd$AMBITO=="Economía" | dd$AMBITO=="Informática" | dd$AMBITO=="Ingeniería y
profesiones afines" | dd$AMBITO=="Arquitectura y construcción" |
dd$AMBITO=="Medicina" | dd$AMBITO=="Enfermería y atención a enfermos")
df$AMBITO <- factor(df$AMBITO)

```



```

#Reagrupamos algunas categorias
library(car)
df$AMBITO <- recode(df$AMBITO, "'Formación de docentes de enseñanza
infantil'='Educación'; 'Formación de docentes de enseñanza
primaria'='Educación'; 'Otra formación de personal docente y ciencias de la
educación'='Educación' ; 'Humanidades'='Humanidades y artes'; 'Artes' =
'Humanidades y artes'; 'Administración y gestión de empresas' = 'Ciencias
económicas y empresariales'; 'Economía'='Ciencias económicas y
empresariales'; 'Otra educación comercial y empresarial'='Ciencias
económicas y empresariales'; ")
dd <- df
rm(df)

# PRACTICAS

PRAC <- c()
dd$PRAC_CR <- as.numeric(dd$PRAC_CR)
dd$EST_XCU <- as.numeric(dd$EST_XCU)

for(i in 1:length(dd$PRAC_CR)){
  if(is.na(dd$PRAC_CR[i]) & is.na(dd$EST_XCU[i])){

    PRAC <- c(PRAC,NA)

  }else if(is.na(dd$PRAC_CR[i]) & is.na(dd$EST_XCU[i])==F){

    PRAC <- c(PRAC, dd$EST_XCU[i])

  }else if(is.na(dd$PRAC_CR[i])==F & is.na(dd$EST_XCU[i])){

    PRAC <- c(PRAC, dd$PRAC_CR[i])

  }else if(dd$PRAC_CR[i]==1 || dd$EST_XCU[i]==1){

    PRAC <- c(PRAC, 1)

  }else if(dd$PRAC_CR[i]==2 & dd$EST_XCU[i]==2){

    PRAC <- c(PRAC, 2)

  }
}

dd <- cbind(dd, PRAC)

dd$PRAC <- as.factor(dd$PRAC)
levels(dd$PRAC) = c("Sí","No", "No aplicable")
summary(dd$PRAC)

# TREBALLA (NO PRACTIQUES)

dd$TRAB_F1 <- as.factor(dd$TRAB_F1)
levels(dd$TRAB_F1) <- c("Sí", "No")

dd$TRAB_F2 <- as.factor(dd$TRAB_F2)
levels(dd$TRAB_F2) <- c(NA, "Sí", "No","No aplicable")

# MEJORA DE CONDICIONES

```

```

dd$BQTRA22 <- as.factor(dd$BQTRA22)
levels(dd$BQTRA22) <- c(NA, "Sí", "No", "No aplicable")

# TREBALL DESPRES D'ACABAR LA CARRERA

dd$HA_TRAB <- as.factor(dd$HA_TRAB)
levels(dd$HA_TRAB) <- c("Sí", "No")

# TEMPS ENTRE FINAL CARRERA I TREBALLAR

dd$BQ_TTRA <- as.factor(dd$BQ_TTRA)
levels(dd$BQ_TTRA) <- c("Continuó al menos 6 meses más en el trabajo que
tenia", "Menos de 3 meses", "De 3 a 6 meses", "De 6 meses a 1 año", "De 1
año a año y medio", "De 1 año y medio a 2 años", "Más de dos años", "No
aplicable")

# Rechazo de algún trabajo (P52)

dd$BSQRE <- as.factor(dd$BSQRE)
levels(dd$BSQRE) <- c(NA, "Sí", "No", "No aplicable")

# Motivos rechazo

dd$BSQREM1 <- as.factor(dd$BSQREM1)
dd$BSQREM2 <- as.factor(dd$BSQREM2)
dd$BSQREM3 <- as.factor(dd$BSQREM3)
dd$BSQREM4 <- as.factor(dd$BSQREM4)
dd$BSQREM5 <- as.factor(dd$BSQREM5)
dd$BSQREM6 <- as.factor(dd$BSQREM6)
dd$BSQREM7 <- as.factor(dd$BSQREM7)

levels(dd$BSQREM1) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQREM2) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQREM3) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQREM4) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQREM5) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQREM6) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQREM7) <- c(NA, "Sí", "No", "No aplicable")

# Maneras de búsqueda primer empleo

dd$BSQCOM1 <- as.factor(dd$BSQCOM1)
dd$BSQCOM2 <- as.factor(dd$BSQCOM2)
dd$BSQCOM3 <- as.factor(dd$BSQCOM3)
dd$BSQCOM4 <- as.factor(dd$BSQCOM4)
dd$BSQCOM5 <- as.factor(dd$BSQCOM5)
dd$BSQCOM6 <- as.factor(dd$BSQCOM6)
dd$BSQCOM7 <- as.factor(dd$BSQCOM7)
dd$BSQCOM8 <- as.factor(dd$BSQCOM8)
dd$BSQCOM9 <- as.factor(dd$BSQCOM9)

levels(dd$BSQCOM1) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM2) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM3) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM4) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM5) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM6) <- c(NA, "Sí", "No", "No aplicable")

```

```

levels(dd$BSQCOM7) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM8) <- c(NA, "Sí", "No", "No aplicable")
levels(dd$BSQCOM9) <- c(NA, "Sí", "No", "No aplicable")

# Dificultades para encontrar trabajo

dd$BSQDIF1 <- as.factor(dd$BSQDIF1)
dd$BSQDIF2 <- as.factor(dd$BSQDIF2)
dd$BSQDIF3 <- as.factor(dd$BSQDIF3)
dd$BSQDIF4 <- as.factor(dd$BSQDIF4)
dd$BSQDIF5 <- as.factor(dd$BSQDIF5)
dd$BSQDIF6 <- as.factor(dd$BSQDIF6)

levels(dd$BSQDIF1) <- c(NA, "Nada importante", "Poco importante", "Bastante
importante", "Muy importante", "No aplicable")
levels(dd$BSQDIF2) <- c(NA, "Nada importante", "Poco importante", "Bastante
importante", "Muy importante", "No aplicable")
levels(dd$BSQDIF3) <- c(NA, "Nada importante", "Poco importante", "Bastante
importante", "Muy importante", "No aplicable")
levels(dd$BSQDIF4) <- c(NA, "Nada importante", "Poco importante", "Bastante
importante", "Muy importante", "No aplicable")
levels(dd$BSQDIF5) <- c(NA, "Nada importante", "Poco importante", "Bastante
importante", "Muy importante", "No aplicable")
levels(dd$BSQDIF6) <- c(NA, "Nada importante", "Poco importante", "Bastante
importante", "Muy importante", "No aplicable")

# Situación profesional primer empleo

dd$PSIT_PR <- as.factor(dd$PSIT_PR)
levels(dd$PSIT_PR) <- c(NA, "Prácticas, formación o becario", "Asalariado
con contrato permanente", "Asalariado con contrato temporal", "Empresario o
trabajador independiente", "Ayuda familiar", "No aplicable")

# Jornada laboral

dd$P_JOR <- as.factor(dd$P_JOR)
levels(dd$P_JOR) <- c(NA, "A tiempo completo", "A tiempo parcial", "No
aplicable")

# Hacia uso de sus conocimientos y habilidades en el primer trabajo

dd$P_CONOC <- as.factor(dd$P_CONOC)
levels(dd$P_CONOC) <- c(NA, "Sí", "No", "No aplicable")

# Situación profesional actual

dd$SIT_PRO <- as.factor(dd$SIT_PRO)
levels(dd$SIT_PRO) <- c(NA, "Prácticas, formación o becario", "Asalariado
con contrato permanente", "Asalariado con contrato temporal", "Empresario o
trabajador independiente", "Ayuda familiar", "No aplicable")

# JORNADA

dd$JORNADA <- as.factor(dd$JORNADA)
levels(dd$JORNADA) <- c(NA, "A tiempo completo", "A tiempo parcial", "No
aplicable")

# Personas que trabajan en la empresa actual

```

```

dd$TR_TAM <- as.factor(dd$TR_TAM)
levels(dd$TR_TAM) <- c(NA, "De 1 a 10 personas", "Entre 11 y 19 personas",
"Entre 20 y 49 personas", "50 o más personas", "No aplicable")

# Empleos distintos desde la obtención del título

dd$BQ_NUM <- as.numeric(dd$BQ_NUM)

# TEMPS TREBALLAT DESDE LA FINALITZACIÓ DE LA CARRERA

dd$BQ_TP_T <- as.factor(dd$BQ_TP_T)
levels(dd$BQ_TP_T) <- c(NA, "Menos de 6 meses", "De 6 meses a 1 año", "De 1
año a año y medio", "De 1 año y medio a 2 años", "2 o más años", "No
aplicable")

# Factores que han influido para conseguir el trabajo

dd$DSTRE41 <- as.factor(dd$DSTRE41)
dd$DSTRE42 <- as.factor(dd$DSTRE42)
dd$DSTRE43 <- as.factor(dd$DSTRE43)
dd$DSTRE44 <- as.factor(dd$DSTRE44)
dd$DSTRE45 <- as.factor(dd$DSTRE45)
dd$DSTRE46 <- as.factor(dd$DSTRE46)

levels(dd$DSTRE41) <- c(NA, "Nada", "Poco", "Algo", "Bastante", "Mucho", "No
aplicable")
levels(dd$DSTRE42) <- c(NA, "Nada", "Poco", "Algo", "Bastante", "Mucho", "No
aplicable")
levels(dd$DSTRE43) <- c(NA, "Nada", "Poco", "Algo", "Bastante", "Mucho", "No
aplicable")
levels(dd$DSTRE44) <- c(NA, "Nada", "Poco", "Algo", "Bastante", "Mucho", "No
aplicable")
levels(dd$DSTRE45) <- c(NA, "Nada", "Poco", "Algo", "Bastante", "Mucho", "No
aplicable")
levels(dd$DSTRE46) <- c(NA, "Nada", "Poco", "Algo", "Bastante", "Mucho", "No
aplicable")

# Nivel de formación más adecuado para realizar el trabajo actual

dd$DISTRE2 <- as.factor(dd$DISTRE2)
levels(dd$DISTRE2) <- c(NA, "Doctorado o postdoctorado", "Título
universitario", "Formación profesional de grado superior", "Formación
profesional de grado medio / bachillerato", "ESO, EGB, etc.", "No aplicable")

```

```

#####
# GRAFICS INSERCIO      #
#####

```

```

library(ggplot2)
library(dplyr)

# Sexe

a <- plyr::count(dd$SEXO)
n <- length(dd$SEXO)

a <- cbind(a, prop=c(round(a[[2]][1]/n*100,2), round(a[[2]][2]/n*100,2)))

count.data <- a %>% arrange(desc(x)) %>% mutate(lab.ypos = cumsum(prop) -
0.5*prop)

mycols <- c("darkgoldenrod2","darkcyan")
pie = ggplot(count.data, aes(x=2, y=prop, fill=x)) + geom_bar(stat =
"identity", color="white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y = lab.ypos, label = prop), color = "white", size=7)+
  scale_fill_manual(values = mycols) + theme_void() + xlim(0.5,2.5)

pie + theme_classic() + theme(axis.line = element_blank(),
                             axis.text = element_blank(),
                             axis.ticks = element_blank(),
                             plot.title = element_text(hjust = 0.5, color
= "gray32")) + labs(x = NULL, y = NULL, fill = "Género")

# RAMA

s <- ggplot(dd, aes(RAMA, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Rama de
estudios",y="Proporción")

# AMBITO

s <- ggplot(dd, aes(AMBITO, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Ámbito de
estudios",y="Proporción")

# AMBITO2

s <- ggplot(dd, aes(AMBITO, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Ámbito de
estudios",y="Proporción")

# Practiques

dd_1 <- subset(dd, PRAC!="No aplicable")

```

```

ggplot(dd_1, aes(PRAC, fill=SEXO)) + geom_bar(aes(PRAC, fill=SEXO,
(..count..)/sum(..count..)), position = "dodge") +
xlab("Prácticas")+ylab("Proporción") + scale_fill_manual("Género", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan"))

# Treballa (no practiques)

ggplot(dd, aes(TRAB_F1, fill=SEXO)) + geom_bar(aes(TRAB_F1, fill=SEXO,
(..count..)/sum(..count..)), position = "dodge") + xlab("Trabaja durante la
carrera")+ylab("Proporción") + ylim(c(0,0.4))+ scale_fill_manual("Género",
values = c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan"))

# Treball relacionat amb els estudis

dd_1 <- subset(dd, TRAB_F2!="No aplicable" & is.na(TRAB_F2)==F)

ggplot(dd_1, aes(TRAB_F2, fill=SEXO)) + geom_bar(aes(TRAB_F2, fill=SEXO,
(..count..)/sum(..count..)), position = "dodge") + xlab("Trabajo relacionado
con los estudios")+ylab("Proporción") + ylim(c(0,0.4)) +
scale_fill_manual("Género", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

#Mejoraron las condiciones ¿? (trabajan en su ambito y continuaron durante
6 meses)

dd_1 <- subset(dd, TRAB_F2 == "Sí" & TRAB_F3=="1" & BQTRA22!="No
aplicable")

ggplot(dd_1, aes(BQTRA22, fill=SEXO)) + geom_bar(aes(BQTRA22, fill=SEXO,
(..count..)/sum(..count..)), position = "dodge") + xlab("Mejoraron las
condiciones")+ylab("Proporción") + ylim(c(0,0.4)) +
scale_fill_manual("Género", values = c("Hombre" = "darkgoldenrod2", "Mujer"
= "darkcyan"))

# Trabajo una vez acabada la carrera

ggplot(dd, aes(HA_TRAB, fill=SEXO)) + geom_bar(aes(HA_TRAB, fill=SEXO,
(..count..)/sum(..count..)), position = "dodge") +
xlab("Trabajo")+ylab("Proporción") + scale_fill_manual("Género", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan")) +ylim(c(0,0.6))

# Tiempo entre la finalización de la carrera y trabajar

dd_1 <- subset(dd, BQ_TTRA!="No aplicable")
dd_1 <- subset(dd_1, BQ_TTRA!="Continuó al menos 6 meses más en el trabajo
que tenia")

s <- ggplot(dd_1, aes(BQ_TTRA, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo", y="Proporción")

```

```

gr4 <- ggplot(dd_1, aes(x=BQ_TTRA, fill=SEXO)) + geom_bar(position="fill")
+ scale_fill_manual("Género", values = c("Hombre" = "darkgoldenrod2",
"Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")
gr4 + coord_flip() + facet_wrap(~ AMBITO)
# Rechazo de algún trabajo (P52)

dd_1 <- subset(dd, is.na(BSQRE)==F & BSQRE != "No aplicable")

ggplot(dd_1, aes(BSQRE, fill=SEXO)) + geom_bar(aes(BSQRE,fill=SEXO,
(..count..)/sum(..count..)),position = "dodge") + xlab("Rechazo
trabajo")+ylab("Proporción") + scale_fill_manual("Género", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan")) + ylim(c(0,0.5))

# Motivos de rechazo (P53)

library(reshape)
library(readxl)
fedreason <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Insercion_laboral/BSQREM
.xlsx")
fedreason<-as.data.frame(fedreason)
fedreason_data <- melt(data=fedreason, id.vars="Motivo")
colnames(fedreason_data)[2]<-"Género"
levels(fedreason_data$Género)[1]<-"Hombre"
levels(fedreason_data$Género)[2]<-"Mujer"

s<-ggplot(fedreason_data, aes(x=Motivo, y=value, fill = Género)) +
  geom_bar(position = "fill", stat = "identity",width=0.9) + coord_flip() +
scale_fill_manual(values=c("darkgoldenrod2","darkcyan"))
s + aes(Motivo,fill=Género) + xlab("Manera") + ylab("Proporción")

### CARACTERÍSTICAS DEL PRIMER TRABAJO

#Maneras de búsqueda

library(reshape)
library(readxl)
fedreason <-
read_excel("C:/Users/alici/OneDrive/Escritorio/TFG/Insercion_laboral/BSQCOM
.xlsx")
fedreason<-as.data.frame(fedreason)
fedreason_data <- melt(data=fedreason, id.vars="Maneras")
colnames(fedreason_data)[2]<-"Género"
levels(fedreason_data$Género)[1]<-"Hombre"
levels(fedreason_data$Género)[2]<-"Mujer"

s<-ggplot(fedreason_data,
aes(x=reorder(Maneras,fedreason_data$Género=="Mujer",sum), y=value, fill =
Género)) +
  geom_bar(position = "fill", stat = "identity",width=0.9) + coord_flip() +
scale_fill_manual(values=c("darkgoldenrod2","darkcyan"))
s + xlab("Manera") + ylab("Proporción")

# Situacion profesional (P58)

dd_1 <- subset(dd, is.na(PSIT_PR)==F & PSIT_PR != "No aplicable")

```

```

s <- ggplot(dd_1, aes(PSIT_PR, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

gr4 <- ggplot(dd_1, aes(x=PSIT_PR, fill=SEXO)) + geom_bar(position="fill")
+ scale_fill_manual("Género", values = c("Hombre" = "darkgoldenrod2",
"Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")
gr4 + coord_flip() + facet_wrap(~ AMBITO)

# Jornada laboral

dd_1 <- subset(dd, is.na(P_JOR)==F & P_JOR!="No aplicable")

ggplot(dd_1, aes(P_JOR, fill=SEXO)) + geom_bar(aes(P_JOR,fill=SEXO,
(..count..)/sum(..count..)),position = "dodge") +
xlab("Jornada")+ylab("Proporción") + scale_fill_manual("Género", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan")) + ylim(c(0,0.4))

s <- ggplot(dd_1, aes(P_JOR, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Jornada",y="Proporción")
+ facet_wrap(~ AMBITO)

# Uso de conocimientos y habilidades

dd_1 <- subset(dd, is.na(P_CONOC)==F & P_CONOC!="No aplicable")

ggplot(dd_1, aes(P_CONOC, fill=SEXO)) + geom_bar(aes(P_CONOC,fill=SEXO,
(..count..)/sum(..count..)),position = "dodge") +
xlab("Jornada")+ylab("Proporción") + scale_fill_manual("Género", values =
c("Hombre" = "darkgoldenrod2", "Mujer" = "darkcyan")) + ylim(c(0,0.5))

### SITUACION LABORAL ACTUAL

# Situación profesional

dd_1 <- subset(dd, SIT_PRO!="No aplicable" & is.na(SIT_PRO)==F)

dd_1$SIT_PRO = factor(dd_1$SIT_PRO, levels = c("No aplicable","Ayuda
familiar","Prácticas, formación o becario","Asalariado con contrato
temporal","Asalariado con contrato permanente","Empresario o trabajador
independiente"))

s <- ggplot(dd_1, aes(x=SIT_PRO, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

## Segun ámbito

s <- ggplot(dd_1, aes(x=SIT_PRO, fill = SEXO)) + geom_bar(position =
"fill")

```



```

s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")
+ facet_wrap(~AMBITO)

# Dimension de la empresa (P75)

dd_1 <- subset(dd, TR_TAM!="No aplicable" & is.na(TR_TAM)==F)

s <- ggplot(dd_1, aes(x=TR_TAM, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

# Tiempo trabajado desde la obtención del título (P78)

dd_1 <- subset(dd, is.na(BQ_TP_T)==F)
dd_1 <- subset(dd_1, BQ_TP_T!="No aplicable")

s <- ggplot(dd_1, aes(BQ_TP_T, fill = SEXO)) + geom_bar(position = "fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

gr4 <- ggplot(dd_1, aes(x=BQ_TP_T, fill=SEXO)) + geom_bar(position="fill")
+ scale_fill_manual("Género", values = c("Hombre" = "darkgoldenrod2",
"Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")
gr4 + coord_flip() + facet_wrap(~ RAMA)

# Factores para conseguir el trabajo

## Conocimientos teoricos

dd_1 <- subset(dd, DSTRE41!="No aplicable" & is.na(DSTRE41)==F)

s <- ggplot(dd_1, aes(x=DSTRE41, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

## Conocimientos practicos

dd_1 <- subset(dd, DSTRE42!="No aplicable" & is.na(DSTRE42)==F)

s <- ggplot(dd_1, aes(x=DSTRE42, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

## Conocimiento de idiomas

dd_1 <- subset(dd, DSTRE42!="No aplicable" & is.na(DSTRE42)==F)

```

```

s <- ggplot(dd_1, aes(x=DSTRE42, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

## Formación o dominio de informática y nuevas tecnologías

dd_1 <- subset(dd, DSTRE42!="No aplicable" & is.na(DSTRE42)==F)

s <- ggplot(dd_1, aes(x=DSTRE42, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

## Manera de ser (personalidad, habilidades sociales, comunicacion,
capacidad de trabajar en grupo)

dd_1 <- subset(dd, DSTRE42!="No aplicable" & is.na(DSTRE42)==F)

s <- ggplot(dd_1, aes(x=DSTRE42, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

## Capacidad de gestión y planificación

dd_1 <- subset(dd, DSTRE42!="No aplicable" & is.na(DSTRE42)==F)

s <- ggplot(dd_1, aes(x=DSTRE42, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Tiempo",y="Proporción")

# Nivel de formación más adecuado para realizar el trabajo actual

dd_1 <- subset(dd, DISTRE2!="No aplicable" & is.na(DISTRE2)==F)

s <- ggplot(dd_1, aes(x=DISTRE2, fill = SEXO)) + geom_bar(position =
"fill")
s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Nivel de
formación",y="Proporción")

s + coord_flip() + theme(plot.title = element_text(hjust =1, vjust = 2,
color="gray32")) + scale_fill_manual("Género", values = c("Hombre" =
"darkgoldenrod2", "Mujer" = "darkcyan")) + labs(x="Nivel de
formación",y="Proporción") + facet_wrap(~AMBITO)

#####
## TESTOS ##
#####

# Practiques

```

```

dd_1 <- subset(dd, PRAC!="No aplicable")

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$PRAC)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Prácticas"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

#Treballa durant la carrera

datos <- data.frame(sexo = dd$SEXO, var = dd$TRAB_F1)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Trabaja"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Treball relacionat amb els estudis

dd_1 <- subset(dd, TRAB_F2!="No aplicable" & is.na(TRAB_F2)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$TRAB_F2)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Relacionado"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Millora condicions

dd_1 <- subset(dd, TRAB_F2 == "Sí" & TRAB_F3=="1" & BQTRA22!="No
aplicable")

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BQTRA22)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Mejora
condiciones"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Treball despres de la carrera

datos <- data.frame(sexo = dd$SEXO, var = dd$HA_TRAB)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Trabajo"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

```

```

# Temps entre acabar la carrera i treballar

dd_1 <- subset(dd, BQ_TTRA!="No aplicable")
dd_1 <- subset(dd_1, BQ_TTRA!="Continuó al menos 6 meses más en el trabajo
que tenia")

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BQ_TTRA)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Tiempo"))
tabla <- tabla[,-1]
tabla <- tabla[,-7]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

d_a <- subset(dd_1, AMBITO=="Medicina")

tabla <- table(d_a$SEXO, d_a$BQ_TTRA, dnn=c("Género", "x"))
tabla <- tabla[,-8]
tabla <- tabla[,-1]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Temps treballant des de l'obtencio del titol

dd_1 <- subset(dd, is.na(BQ_TP_T)==F)
dd_1 <- subset(dd_1, BQ_TP_T!="No aplicable")

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BQ_TP_T)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Tiempo"))
tabla <- tabla[,-6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

d_a <- subset(dd_1, RAMA=="Ciencias de la salud")

tabla <- table(d_a$SEXO, d_a$BQ_TP_T, dnn=c("Género", "Ciencias de la
salud"))
tabla <- tabla[,-6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Rechazo trabajo

dd_1 <- subset(dd, is.na(BSQRE)==F & BSQRE != "No aplicable")
datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BSQRE)

```

```

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Rechazo"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

## Motivos rechazo

dd_1 <- subset(dd, BSQRE=="Sí" & is.na(BSQREM7)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BSQREM7)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Rechazo"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Maneras de busqueda en su primer empleo

dd_1 <- subset(dd, BSQCOM9!="No aplicable" & is.na(BSQCOM9)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BSQCOM9)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Búsqueda"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Dificultades para encontrar primer empleo

dd_1 <- subset(dd, BSQDIF1!="No aplicable" & is.na(BSQDIF1)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$BSQDIF1)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Dificultades"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Situacion profesional primer empleo

dd_1 <- subset(dd, is.na(PSIT_PR)==F & PSIT_PR != "No aplicable")

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$PSIT_PR)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Situación
profesional"))
tabla <- tabla[, -6]
chisq.test(x=tabla)
chisq.test(x=tabla)$residuals

```

```

chisq.test(x=tabla)$stdres

# Jornada laboral primer empleo

dd_1 <- subset(dd, is.na(P_JOR)==F & P_JOR!="No aplicable")

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$P_JOR)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Jornada"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

## Jornada segun ambito

d <- subset(dd_1, AMBITO=="Medicina")

datos <- data.frame(sexo = d$SEXO, var = d$P_JOR)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Jornada"))
tabla <- tabla[, -3]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

prop.table(table(dd_1$AMBITO, dd_1$SEXO), 1) #proporción de hombres y
mujeres en cada ámbito

# AMBITO

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$AMBITO)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Ambito"))

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Situación profesional empleo actual

dd_1 <- subset(dd, SIT_PRO!="No aplicable" & is.na(SIT_PRO)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$SIT_PRO)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Situación
profesional"))
tabla <- tabla[, -6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

## Segun ambito

d <- subset(dd_1, AMBITO=="Medicina")

```

```

datos <- data.frame(sexo = d$SEXO, var = d$SIT_PRO)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Situación
profesional"))
tabla <- tabla[,-1]
tabla <- tabla[,-1]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Dimension de la empresa actual

dd_1 <- subset(dd, TR_TAM!="No aplicable" & is.na(TR_TAM)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$TR_TAM)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Dimensión empresa"))
tabla <- tabla[,-5]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Factores que han influido en la obtencion del trabajo

dd_1 <- subset(dd, DSTRE46!="No aplicable" & is.na(DSTRE46)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$DSTRE42)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Factores"))
tabla <- tabla[,-6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

# Nivel de formación más adecuado para realizar el trabajo actual

dd_1 <- subset(dd, DISTRE2!="No aplicable" & is.na(DISTRE2)==F)

datos <- data.frame(sexo = dd_1$SEXO, var = dd_1$DISTRE2)

tabla <- table(datos$sexo, datos$var, dnn=c("Género", "Nivel de
formacion"))
tabla <- tabla[,-6]

chisq.test(x=tabla)
chisq.test(x=tabla)$residuals
chisq.test(x=tabla)$stdres

#####
## ASSOCIATION RULES ##

```

```
#####

library(arules)
library("arulesViz")

## BLOQUE 1

d_ass <- subset(dd, select = c(SEXO, AMBITO, PRAC, TRAB_F1, TRAB_F2,
BQTRA22))
d_ass <- subset(d_ass, is.na(PRAC)==F & is.na(TRAB_F2)==F &
is.na(BQTRA22)==F)

dtrans <- as(d_ass, "transactions")

inspect(head(dtrans,10))
summary(dtrans)
itemFrequencyPlot(dtrans, support=0.1, cex.names = 1,ylim=c(0,1.2))
itemFrequencyPlot(dtrans, support=0.05, cex.names = 1,ylim=c(0,1.0))

rulesDtrans <- apriori(dtrans, parameter = list(support = 0.15, confidence
= 0.5 , minlen=2, maxlen=4))
summary(rulesDtrans)

inspect(rulesDtrans)
plot(rulesDtrans, measure = c("support", "lift"), shading = "confidence")
#order es el numero de items qe hay
plot(rulesDtrans, method = "two-key plot")
plot(rulesDtrans, method = "grouped")
plot(rulesDtrans, method = "paracoord")
plot(rulesDtrans, method = "graph", control=list(type="items"))

summary(dtrans)
image(dtrans)
dim(dtrans)
image(dtrans[sample(1:length(dtrans),20)])

#Filtramos las reglas, quedandonos con aquellas donde aparezca el género y
eliminando las que tengan no aplicables

rules_sub_h<-subset(rulesDtrans,subset= ((rhs %in% "SEXO=Hombre" & lift > 1
& !( rhs %pin% "=No aplicable")& !( lhs %pin% "=No aplicable")) | (lhs %in%
"SEXO=Hombre" & lift > 1 & !( rhs %pin% "=No aplicable")& !( lhs %pin% "=No
aplicable"))))
rules_sub_m<-subset(rulesDtrans,subset= ((rhs %in% "SEXO=Mujer" & lift > 1
& !( rhs %pin% "=No aplicable")& !( lhs %pin% "=No aplicable")) | (lhs %in%
"SEXO=Mujer" & lift > 1 & !( rhs %pin% "=No aplicable")& !( lhs %pin% "=No
aplicable"))))

inspect(rules_sub_h)
inspect(rules_sub_m)

rules_hm<- c(rules_sub_h,rules_sub_m)
```



```

#Reglas juntas

inspect(rules_hm)

#Buscamos reglas redundantes
subset.matrix <- is.subset(rules_hm, rules_hm, sparse=F)
subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
redundant <- colSums(subset.matrix, na.rm=T) >= 1
which(redundant)
inspect(rules_hm[!redundant])
inspect(rules_hm[])

# Eliminamos reglas redundantes
rules.pruned <- rules_hm[!redundant]
rules.pruned <- sort(rules.pruned, by="lift")
inspect(rules.pruned)

# BLOQUE TOTAL

library(RColorBrewer)
palette(brewer.pal(n = 4, name = "Set2"))

d_ass <- subset(dd, select = c(SEXO, AMBITO, PRAC, TRAB_F1, TRAB_F2,
BQTRA22, BQ_TTRA, BSQRE, PSIT_PR, P_CONOC, SIT_PRO, TR_TAM, BQ_TP_T))
d_ass <- subset(d_ass, is.na(PRAC)==F & is.na(TRAB_F2)==F &
is.na(BQTRA22)==F & is.na(BQ_TTRA)==F & is.na(BSQRE)==F & is.na(PSIT_PR)==F
& is.na(P_CONOC)==F & is.na(SIT_PRO)==F & is.na(TR_TAM)==F &
is.na(BQ_TP_T)==F)

dtrans <- as(d_ass, "transactions")

inspect(head(dtrans, 10))
summary(dtrans)

rulesDtrans <- apriori(dtrans, parameter = list(confidence = 0.4 ,
minlen=2))
summary(rulesDtrans)

plot(rulesDtrans, measure = c("support", "confidence"), shading =
"lift", col=c(8:1))

rulesDtrans <- apriori(dtrans, parameter =
list(support=0.1, confidence=0.4, minlen=2, maxlen=3))
summary(rulesDtrans)

rules_sub_h <- subset(rulesDtrans, subset= ((rhs %in% "SEXO=Hombre" &
support>0.15 & confidence>0.5 & lift > 1 & !(rhs %pin% "=No aplicable") &
!(lhs %pin% "=No aplicable")) | (lhs %in% "SEXO=Hombre" & support>0.15 &
confidence>0.5 & lift > 1 & !(rhs %pin% "=No aplicable") & !(lhs %pin%
"=No aplicable"))))
rules_sub_m <- subset(rulesDtrans, subset= ((rhs %in% "SEXO=Mujer" &
support>0.25 & confidence>0.5 & lift > 1 & !(rhs %pin% "=No aplicable") &

```

```

!( lhs %pin% "=No aplicable")) | (lhs %in% "SEXO=Mujer" & support>0.25 &
confidence>0.5 & lift > 1 & !( rhs %pin% "=No aplicable")& !( lhs %pin%
"No aplicable"))))

inspect(rules_sub_h)
inspect(rules_sub_m)

rules_hm2<- c(rules_sub_h,rules_sub_m)

inspect(rules_hm2)

ruleshm_r<-subset(rules_hm2,subset= !(duplicated(rules_hm2@quality$count)))

inspect(ruleshm_r)

# BLOQUE TOTAL 2

d_prueba <- subset(dd, select = c(SEXO, AMBITO, PRAC, TRAB_F1, TRAB_F2,
BQTRA22,BQ_TTRA,BSQRE,BSQREM1,BSQREM2,BSQREM3,BSQREM4,BSQREM5,BSQREM6,BSQRE
M7,BSQCOM1,BSQCOM2,BSQCOM3,BSQCOM4,BSQCOM5,BSQCOM6,BSQCOM7,BSQCOM8,BSQCOM9,
BSQDIF1,BSQDIF2,BSQDIF3,BSQDIF4,BSQDIF5,BSQDIF6,PSIT_PR,P_JOR,P_CONOC,SIT_P
RO,JORNADA,TR_TAM,
BQ_TP_T,DSTRE41,DSTRE42,DSTRE43,DSTRE44,DSTRE45,DSTRE46))
#d_prueba <- subset(d_ass, is.na(PRAC)==F & is.na(TRAB_F2)==F &
is.na(BQTRA22)==F & is.na(BQ_TTRA)==F & is.na(BSQRE)==F & is.na(PSIT_PR)==F
& is.na(P_CONOC)==F & is.na(SIT_PRO)==F & is.na(TR_TAM)==F &
is.na(BQ_TP_T)==F)

dtrans <- as(d_prueba, "transactions")
rulesDtrans <- apriori(dtrans, parameter =
list(support=0.1,confidence=0.4,minlen=2, maxlen=2))
summary(rulesDtrans)

rules_sub_h<-subset(rulesDtrans,subset= ((rhs %in% "SEXO=Hombre" &
support>0.15 & confidence>0.5 & lift > 1 & !( rhs %pin% "=No aplicable")&
!( lhs %pin% "=No aplicable")) | (lhs %in% "SEXO=Hombre" & support>0.15 &
confidence>0.5 & lift > 1 & !( rhs %pin% "=No aplicable")& !( lhs %pin%
"No aplicable"))))
rules_sub_m<-subset(rulesDtrans,subset= ((rhs %in% "SEXO=Mujer" &
support>0.25 & confidence>0.5 & lift > 1 & !( rhs %pin% "=No aplicable")&
!( lhs %pin% "=No aplicable")) | (lhs %in% "SEXO=Mujer" & support>0.25 &
confidence>0.5 & lift > 1 & !( rhs %pin% "=No aplicable")& !( lhs %pin%
"No aplicable"))))

inspect(rules_sub_h)
inspect(rules_sub_m)

rules_hm2<- c(rules_sub_h,rules_sub_m)

inspect(rules_hm2)

ruleshm_r<-subset(rules_hm2,subset= !(duplicated(rules_hm2@quality$count)))

inspect(ruleshm_r)

#####

```

```
## DADES TWITTER ##
#####

library(twitter)
library(stringr)
library(tidytext)
library(dplyr)
library(broom)
library(tidyr)
library(ggplot2)
library(quanteda)
library(purrr)
library(wordcloud)
library(RColorBrewer)
library(tm)

consumer_key <- "Fa3Xdrm00uAzGR8xYSXT7LAhc"
consumer_secret <- "bHQQ2J4Ce2Qp2AoDX0isTWBS6r4EIMfXMCrXlpw5dP7N5dZjBc"

access_token <- "421813052-iEiEhRCH0HfrrpYvQc6dWBYShtdycLIECG3YKBwnq"
access_secret <- "JToaUkKoqOzDMtFrzKZ0GL5GV6NYE81C9mwD7gd9QIN6n"

setup_twitter_oauth(consumer_key, consumer_secret, access_token,
access_secret)

tw_1 = searchTwitter("#BrechaSalarial", n=1000)

tw_2 = searchTwitter("#Igualdad", n=10000, since = '2018-06-11')

tw_3 = searchTwitter("#InsercionLaboral", n=1000, since = '2018-06-11')

d1 = twListToDF(tw_1)
d2 = twListToDF(tw_2)
d3 = twListToDF(tw_3)

# Seleccio variables

tweets1 <- d1 %>% select(text)
tweets2 <- d2 %>% select(text)
tweets3 <- d3 %>% select(text)

# Neteja del text

limpiar_tokenizar <- function(texto){
  # El orden de la limpieza no es arbitrario
  # Se convierte todo el texto a minúsculas
  nuevo_texto <- tolower(texto)
  # Eliminación de páginas web (palabras que empiezan por "http." seguidas
  # de cualquier cosa que no sea un espacio)
  nuevo_texto <- str_replace_all(nuevo_texto,"http\\S*", "")
  # Eliminación de signos de puntuación
  nuevo_texto <- str_replace_all(nuevo_texto,"[:punct:]", " ")
  # Eliminación de números
  nuevo_texto <- str_replace_all(nuevo_texto,"[:digit:]", " ")
  # Eliminación de espacios en blanco múltiples
  nuevo_texto <- str_replace_all(nuevo_texto,"[\\s]+", " ")
  # Tokenización por palabras individuales
```

```

nuevo_texto <- str_split(nuevo_texto, " ")[[1]]
# Eliminación de tokens con una longitud < 2
nuevo_texto <- keep(.x = nuevo_texto, .p = function(x){str_length(x) >
1})
return(nuevo_texto)
}

tweets1 <- tweets1 %>% mutate(texto_tokenizado = map(.x = text, .f =
limpiar_tokenizar))
tweets2 <- tweets2 %>% mutate(texto_tokenizado = map(.x = text, .f =
limpiar_tokenizar))
tweets3 <- tweets3 %>% mutate(texto_tokenizado = map(.x = text, .f =
limpiar_tokenizar))

tweets_tidy1 <- tweets1 %>% select(-text) %>% unnest()
tweets_tidy1 <- tweets_tidy1 %>% rename(token = texto_tokenizado)

tweets_tidy2 <- tweets2 %>% select(-text) %>% unnest()
tweets_tidy2 <- tweets_tidy2 %>% rename(token = texto_tokenizado)

tweets_tidy3 <- tweets3 %>% select(-text) %>% unnest()
tweets_tidy3 <- tweets_tidy3 %>% rename(token = texto_tokenizado)

lista_stopwords <- c('el', 'la',
'de', 'en', 'los', 'para', 'con', 'su', 'las', 'más', 'es', 'se', 'rt', 'que', 'del', 'a
l', 'una', 'lo', 'sobre', 'no', 'una', 'me', 'te', 'mi', 'un', 'sus', 'han', 'yo', 'ha',
'mx', 'unas', 'por')

tweets_tidy1 <- tweets_tidy1 %>% filter(!(token %in% lista_stopwords))
tweets_tidy2 <- tweets_tidy2 %>% filter(!(token %in% lista_stopwords))
tweets_tidy3 <- tweets_tidy3 %>% filter(!(token %in% lista_stopwords))

tweets_brecha <- tweets_tidy1
tweets_igualdad <- tweets_tidy2
tweets_insercion <- tweets_tidy3

# Wordcloud (brecha salarial)

df1 <- tweets_brecha %>% count(token)

wordcloud(words = df1$token, freq = df1$n,
max.words = 350, random.order = FALSE, rot.per = 0.35,
colors = brewer.pal(8, "Dark2"))

#Wordcloud2 (igualdad)

df2 <- tweets_igualdad %>% count(token)

wordcloud(words = df2$token, freq = df2$n,
max.words = 350, random.order = FALSE, rot.per = 0.35,
colors = brewer.pal(8, "Dark2"))

#Wordcloud3 (insercion)

df3 <- tweets_insercion %>% count(token)

```

```

wordcloud(words = df3$token, freq = df3$n,
          max.words = 400, random.order = FALSE, rot.per = 0.35,
          colors = brewer.pal(8, "Dark2"))

#####
## ANALISIS SERIES TEMPORALES GOOGLE TRENDS ##
#####

library(ggplot2)
devtools::install_github("brisneve/ggplottimeseries")
library(ggplottimeseries)
library(tidyr)
library(pastecs)
library(tseries)

brecha <- read.csv(file="brecha_salarial.csv", header = T, sep = ",")
names(brecha) <- c("semana","brecha_salarial")
brecha$semana <- as.Date(brecha$semana)
brecha$brecha_salarial <- as.numeric(brecha$brecha_salarial)

insercion <- read.csv(file="insercion_laboral.csv", header = T, sep = ",")
names(insercion) <- c("semana","insercion_laboral")
insercion$semana <- as.Date(insercion$semana)
insercion$insercion_laboral <- as.numeric(insercion$insercion_laboral)

igualdad <- read.csv(file="igualdad.csv", header = T, sep = ",")
names(igualdad) <- c("semana","igualdad_t")
igualdad$semana <- as.Date(igualdad$semana)
igualdad$igualdad_t <- as.numeric(igualdad$igualdad_t)

ts_brecha <- ts(brecha$brecha_salarial, start=c(2014), frequency = 48)
ts_insercion <- ts(insercion$insercion_laboral, start = c(2014), frequency
= 48)
ts_igualdad <- ts(igualdad$igualdad_t, start = c(2014), frequency = 48)

todo <- read.csv(file="igualdad_brecha_insercion.csv", header = T, sep =
",")
names(todo) <- c("semana","igualdad","brecha","insercion")
todo$semana <- as.Date(todo$semana)
todo$igualdad <- as.numeric(todo$igualdad)
todo$brecha <- as.numeric(todo$brecha)
todo$insercion <- as.numeric(todo$insercion)

#####
#Brecha#
#####

ggplot(data = brecha, aes(x = semana, y = brecha_salarial)) +
  geom_line(color = "darkslategrey",size=1.5) + ylab("Interés a lo largo
del tiempo") +xlab("Tiempo")

## Test de tendencia

trend.test(ts_brecha, R=1) #Trend if p-value <0.05

```

```

##Descomposicion

df <- dts1(brecha$semana,brecha$brecha_salarial,48, type = "additive")
#plots decomposed time series into one figure
ggdecompose(df)+
  xlab("Tiempo")+
  ylab("Interés a lo largo del tiempo")+geom_line(colour="darkslategrey")

#####
#Insercion#
#####

ggplot(data = insercion, aes(x = semana, y = insercion_laboral)) +
  geom_line(color = "darkslategrey",size=1.5)+ ylab("Interés a lo largo del
tiempo")+xlab("Tiempo")

## Test de tendencia

trend.test(ts_insercion, R=1)

## Descomposicion

df <- dts1(insercion$semana,insercion$insercion_laboral,48, type =
"additive")
#plots decomposed time series into one figure
ggdecompose(df)+
  xlab("Tiempo")+
  ylab("Interés a lo largo del tiempo")+geom_line(colour="darkslategrey")

#####
#Igualdad#
#####

ggplot(data = igualdad, aes(x = semana, y = igualdad_t)) +
  geom_line(color = "darkslategrey",size=1.5)+ ylab("Interés a lo largo del
tiempo")+xlab("Tiempo")

## Test de tendencia

trend.test(ts_igualdad, R=1)

## Descomposicion

df <- dts1(igualdad$semana,igualdad$igualdad_t,48, type = "additive")
#plots decomposed time series into one figure
ggdecompose(df)+
  xlab("Tiempo")+
  ylab("Interés a lo largo del tiempo")+geom_line(colour="darkslategrey")

#####
#TOTAL#
#####

```

```

ggplot(data = todo, aes(x = semana)) +
  geom_line(aes(y=igualdad),color = "turquoise4",size=1) +
  geom_line(aes(y=brecha),color = "darkorchid3",size=1) +
  geom_line(aes(y=insercion),color = "peru",size=1) +
  ylab("Interés a lo largo del tiempo")+xlab("Tiempo") + legend()

library(tidyverse)
df <- todo %>% select(semana, igualdad, brecha, insercion) %>% gather(key =
"variable", value = "value", -semana)

ggplot(df, aes(x = semana, y = value)) +
  geom_line(aes(color = variable), size=1) +
  scale_color_manual(values = c("darkorchid3", "turquoise4", "peru")) +
  ylab("Interés a lo largo del tiempo") + xlab("Tiempo")

```

Encuestas